

Kyoto challenge has just begun

A move by Russia to support the Kyoto Protocol should usher in an era of international collaboration in mitigating climate change. Validating emissions trading and bringing developing economies into the fold are the next priorities.

Russia's climate politics are something of a mystery to outsiders, but the Russian government has reportedly decided to ratify the Kyoto Protocol on climate change, and approval by the parliament should be a formality. Thus the last obstacle to the protocol's coming into force is set to be removed.

The decision is emerging in a way that lacks conviction from Russia's political and scientific élite (see *Nature* **431**, 12–13; 2004), and the country's Kyoto sceptics can still delay matters. What is important, however, is that the first multilaterally binding climate-protection regime now seems certain to see the light of day.

Even more important is the need to tackle the treaty's limitations. The legal force behind some of its key rules — including penalties for countries emitting more greenhouse gases than they should — is questionable. And excess emitters have little to fear: compensation for emissions at a later date could be endlessly postponed.

Another penalty is suspension from emissions trading. But whether the creation of an international emissions market — essential to reduce the costs of implementing the protocol — will provide an efficient tool for reducing 'hot air' is yet to be seen. It will soon be tested in the European Union (EU), where emissions trading is set to begin in January. The EU's political leaders, who have pressed ahead despite reservations from large European industries, would be delighted if the Kyoto Protocol were to come into effect then.

Many details of the treaty still await clarification. But its true significance is its potential to establish confidence in the practicability of a complex international climate-protection agreement. In particular, Russia's participation will greatly increase the scope for buying and selling emissions rights, and for gaining credits for exporting 'clean development' technologies — key issues for European, Canadian and Japanese industries concerned about the fairness and liquidity of the international emissions market. Whether emissions can be checked against permitted levels remains a key technical challenge.

Russia's ratification should provide a push towards future climate negotiations, and may even prompt the next US administration to take a constructive role. And the authority and credibility of the International Panel on Climate Change can only benefit as well.

The problem of global warming is here to stay, however. Fossil fuels still account for some 90% of the world's energy consumption and are still in abundant supply. Hundreds of millions of people in poorer countries have more spending power, and their consumption is surging, pushing up their energy demand. Any emissions control strategy is therefore ultimately doomed to fail without the inclusion of tomorrow's mega-economies, which are exempted from the need to cut emissions from 1990 levels. Russia's wobbly goodwill provides a glimmer of hope, but our planet's future climate will be determined, above all, in China and India. ■

Futures of artificial life

Researchers involved in synthetic biology need to take steps to engage more with the public.

In 1975, when genetic engineering was still young, the leaders in the field called a meeting at Asilomar, a seaside conference centre in California, where they thrashed out the possible environmental and health risks of the powerful new gene-splicing techniques that they were wielding. They not only agreed important containment guidelines for certain kinds of work, but achieved something potentially more valuable: the wide press coverage they received won the public's trust that scientists were behaving responsibly.

Today that trust is on shaky ground. Controversies over genetically engineered crops and embryo research are leading people to question how carefully scientists consider the possible consequences of their work before barrelling ahead. This is no small concern for science, as it has already led to restrictions.

At the same time, biologists have come to feel increasingly secure in the belief that some ecological nightmare is not likely to spring out of a graduate student's Petri dish. Every day for decades they have been transferring modified genes into microbes, nematodes and mice. At least some of the results — the errant fruitfly or the culture tube spilled in the sink — have no doubt escaped into the environment, without producing a biological Chernobyl.

Is that confidence in step with the technology? The tools now available to the molecular biologist have the potential to provide a stunning array of benefits, for both biomedicine and basic biology. Researchers are learning to understand and manipulate the genetic

circuits that control cells. They can transfer entire synthetic pathways to bacteria to make drugs that must otherwise be extracted from rare plants at great cost. Viral genomes can be synthesized chemically in weeks, and bacterial genomes will soon be within reach.

Through such technologies, a new field of synthetic biology is emerging (see page 624). Bacteria and yeast have been engineered to build proteins impossible in nature, and with novel properties, by the addition of synthetic amino acids. Several groups are even working on assembling simple cells from basic components. This is no longer a matter just of moving genes around. This is shaping life like clay.

Members of the synthetic-biology community have begun to discuss the possible risks, and ethical implications, of their work. But there is no plan as yet for anything like another Asilomar. In one sense, it may be too soon. The scope of these tools is much broader than that of recombinant DNA, and it is certain to be more difficult to foresee what the actual risks are.

But perhaps such discussions can't come soon enough. What will happen if biologists announce that they have made the first living cells from scratch without having demonstrated to the public any concern for the implications? Researchers must do more than talk among themselves. They must demonstrate publicly that they are willing to consult and reflect carefully about risk — both perceived and genuine — and to moderate their actions accordingly. The need for public trust, significant in 1975, is all the greater today. ■





Smell of success
Medicine Nobel celebrates a common sense
p616



Chart attack
Sceptics run down supercomputer top 500 ranking
p618



New dawn
Britain gives green light for Aurora mission
p619



Pipe dreams
Environmentalists float plan to drain flooded canyon
p620

US universities up in arms over licence plans for foreign staff

Geoff Brumfiel, Washington

Scientists and students from countries such as China and India may soon be required to obtain special licences before they can operate laboratory equipment in the US universities where they work.

The change would require scientists from a list of "countries of concern" — including Russia, Pakistan, India and China — to seek a licence from the US Department of Commerce before operating, installing, maintaining or repairing equipment on the department's extensive list of controlled technologies. It is being proposed by the department as a "clarification" of an existing regulation.

But the idea is raising alarm among top university officials. They fear that the modification will at best create a mountain of additional paperwork, and at worst restrict international scientists' ability to participate in research. A group of 22 university presidents, led by Charles Vest of the Massachusetts Institute of Technology in Cambridge, wrote to government officials on 9 September to express their concern. Their letter to Condoleezza Rice — President George Bush's national security adviser — and other top officials warned that the proposed changes "would compromise the international competitiveness and leadership of our institutions".

At the heart of the debate is a long-standing set of regulations covering the export of sensitive technologies to foreign countries. Known broadly as the Export Administration Regulations, the rules are designed to prevent the transfer of equipment that could be used to further a nation's military capabilities. The regulations also prohibit citizens from these countries from obtaining detailed knowledge of controlled technologies while visiting the United States. Foreign nationals from countries of concern must obtain a licence from

Commerce Control List - Index

Description

Avionic equipment, parts, and components	
Avionics EMP/EMI protection technology	
Bacillus anthracis	
Bacteria	
Bacteria	
Bacteria	

Commerce Control List - Index

Description

Airborne communication equipment	
Airborne radar equipment	
Aircraft	
Aircraft, civil	
Aircraft, demilitarized	
Aircraft, n.e.s.	
Aircraft, trainer	
Aircraft, transport	
Aircraft, military	
Aircraft, intermediate	
Aircraft, parts	
Alexandria	
Alexandria	
Align & c	
Alignment	
Alkylphenol	
All flash	
Alloy strip	
Alloyed	
Alloyed	

Commerce Control List - Index

Description

Blanks, Zinc selenide (ZnSe) substrate	
Blanks, Zinc sulphide (ZnS) substrate	
Blowers, axial flow/centrifugal/positive displacement/turbo	
Blowers, positive displacement/centrifugal/axial flow, gas	
Bluetongue virus	
Boats	
Body armor	
Boilers, Marine	
Bombs, smoke	
Bonding equipment	
Boring equipment, large	
Boring machines (CNC)	
Boring machines (CNC)	
Boron	

On a roll: there are thousands of items on a US list of technologies that may require licences for their use by foreigners in academic labs.

the commerce department's Bureau of Industry and Security before working on technologically sensitive equipment.

Historically, universities and government laboratories doing basic research have been exempt from these rules. But in a report issued in March, the commerce department's independent inspector-general warned that although research results may be exempt, the equipment used to conduct research is not, and that researchers should be required to apply for licences before using it.

"We think that the implications are extremely serious," says Robert Hardy, director of contracts and intellectual property management at the Council on Government Relations, a Washington-based association for research universities. "It's one thing to place restrictions on who can come into the country to work and study. But what is being proposed now are further controls on what they're allowed to do on campus."

Hardy predicts that merely determining

which pieces of lab equipment are considered as "controlled" will prove a formidable task. The list includes thousands of entries, ranging from those with an obvious security application, such as "body armor", to more general items such as "bacteria".

"I think it is fair to say that most universities don't even know what they have," Hardy says. Even if the controlled equipment can be inventoried, he says, filing individual licences for each foreign student and researcher working on the apparatus would place enormous strain on university resources.

Peter Lichtenbaum, assistant secretary for export administration at the commerce department, says he is aware of the problems that the revised regulations could pose, and he is working with universities to address the issue.

Lichtenbaum stresses that the legal understanding of what constitutes the "operation" of a piece of equipment is different from the usual definition. Pressing a few buttons on a machine to make a measurement is unlikely to fall under the regulations, he says. But a licence would be required if technologically significant information was gained from using the equipment. He also says the department is considering an exemption that would allow foreign researchers access to the minimum technology needed to operate equipment safely and effectively.

But he says that universities and foreign scientists must be prepared to accept some changes in the way they do research. "We know that foreign governments are targeting universities," he says. "It's not something we can turn a blind eye to. This will require some work and there's no way around that."

The commerce department is expected to issue a draft regulation later this autumn, with a final version to be published next year. ■

Science of smell wins medicine Nobel

Alison Abbott

How would you react if you got that call from Stockholm? Linda Buck, a neuroscientist from Seattle, fumbled her phone in the dark and accidentally hung up. It was 2.30 in the morning, mind you. Co-winner Richard Axel, her former lab chief at Columbia University in New York, managed to take his call without mishap.

The pair won the 2004 Nobel Prize in Physiology or Medicine this week for their seminal work on olfaction — the sensory system concerned with smell. Smell is crucial to animals: they use it to recognize other animals, identify territories and seek out mates, for example. Humans rely on it rather less, but still benefit from being able to smell burning from a distance, for instance, or tell if meat has gone bad.

Twenty years ago, olfaction was the Cinderella of sensory physiology. Back then, neuroscientists were more enamoured of the visual system — and were making great progress towards cracking its mysteries. By 1986, the three main colour receptors — for red, blue and green light — had all been discovered, and scientists had a clear idea of the combinatory code that allows us to perceive light of different wavelengths and identify all the colours of the rainbow.

The small community of olfactory scientists were still working in the dark, however. It was understood that the cells in the lining of the nose bind the molecules that comprise odours, then send electrical signals to the olfactory bulb in the brain, and that from there the information must be relayed to higher parts of the brain concerned with smell recognition. But researchers were confined to sticking electrodes into brains without even knowing what receptors the odour molecules bind to.

Persistent effort

Buck and Axel changed all that. As a post-doc at Axel's Howard Hughes Medical Institute (HHMI) laboratory at Columbia University, Buck became fixated with the problem of the mysterious olfactory receptors. Picking up on recent hints in the scientific literature that smell receptors could be related to the small family of vision receptors, she adapted a freshly developed tool, the polymerase chain reaction (PCR), to serve her ends. PCR amplifies specific genes to detectable levels, and Buck used it in an effort to flush out her receptors in rats. This approach was novel at the time, but has since become routine. After initially failing, she modified the approach — and failed again. And again. Then, after six years, she succeeded.

Together with Axel, she published a paper that shook the world of sensory physiology



Perfect sense: Richard Axel and Linda Buck discovered the receptors that bind odour molecules.

(see *Cell* 65, 175–187; 1991). The receptors were indeed part of the same family as visual receptors — a family that signals through a class of protein known as G proteins. But the olfactory sub-family wasn't the anticipated tight-knit group: it had about a thousand members. Similarly unexpected was the finding that each cell in the lining of the nose contains only one type of olfactory receptor. Each of these cells sends projections to the olfactory bulb. The two researchers went on to show independently that all cells bearing a particular receptor converge on a precise region in the olfactory bulb.

The next big advance will be to understand exactly what happens next — how the signal becomes the perception of a smell with all its associations — the danger of fire, for example, or the hope of love in a perfume. Scientists believe that olfaction will serve as a model for understanding how our brain processes all the sensory information gleaned from our world.

Both Axel and Buck are still working on the higher processing of olfactory signals, but the field is much more crowded now. Their seminal 1991 paper drew researchers into the field in droves. Among them was Lawrence Katz, a self-confessed “hard-core vision scientist” at Duke University Medical Center in Durham, North Carolina. Katz says that Buck and Axel's work “had a tremendous influence on how I thought about how sensory systems could be organized in the brain”.

Like many of his colleagues, Katz soon became convinced that the olfactory system was much more likely to help illumi-

nate the general problem of sensory perception than was vision. “The perception of the olfactory signal — smell — is probably going to be much more closely linked to the molecular signal itself, in contrast to visual perception, whose complexities require much more processing of the signal input,” he says.

Colleagues ascribe Axel and Buck's success in unearthing the roots of olfaction to their personal doggedness, underpinned by the steady, powerful support that the HHMI gives its researchers. “It would have been hard to do this if you were required to produce regular publications to support your next grant,” points out Stuart Firestein, a neuroscientist at Columbia University in New York, who also works on olfaction. The HHMI has nurtured 13 Nobel prizewinners since it was established in 1984.

Firestein says that the award is also a shot in the arm for basic research, because there is, as yet, no direct medical application for Axel and Buck's findings. He notes, nonetheless, that half of all drugs work through G-protein receptors of different types. So having another thousand to investigate as a result of their approach “is certain to help medicine indirectly”, he says.

Buck becomes only the seventh woman to win the Nobel prize for physiology or medicine; fewer than one in 25 laureates in this category have been women. ■

For more news and analysis go to
news @ nature.com
www.nature.com/news

Colourful work on quarks scoops triple crown

Philip Ball

While the Nobel prize for medicine has been awarded for work on smell, this year's physics Nobel stems from research on colour.

But it is not exactly colour as we know it. The prize has been scooped for work on 'colour interaction': the poetic way in which physicists refer to the strong nuclear force, which binds together the fundamental particles called quarks that make up protons and neutrons in an atomic nucleus.

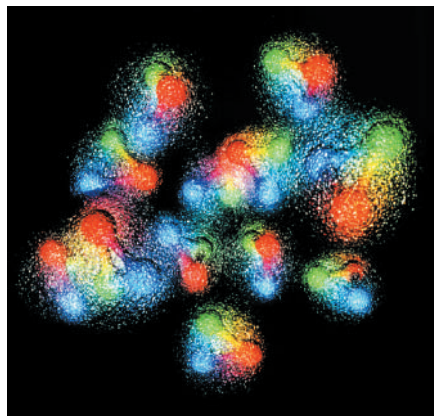
The prize is shared by David Gross of the University of California, Santa Barbara, David Politzer of the California Institute of Technology and Frank Wilczek of the Massachusetts Institute of Technology for their work illuminating the bizarre nature of this force — and making it calculable, at least some of the time.

"It's fantastic news," says particle physicist Jeff Forshaw of the University of Manchester, UK. "Until they came along, we were really struggling to understand the strong force."

Of the four basic forces in nature — gravity, electromagnetism, and the weak and strong nuclear forces — the strong force is particularly counter-intuitive, with the highly unusual property of getting stronger as the quarks get further apart.

This feature, comparable to the behaviour of a rubber band as it is stretched, explains why quarks never appear on their own, but always in combinations. Protons and neutrons each contain a triplet of quarks.

In 1973, Gross, with his graduate student Wilczek, and Politzer published two papers back to back in which they outlined the way



Strong theory: Nobel winners cast light on the gluons that hold nuclear particles together.

the strong force operated (D. J. Gross and F. Wilczek *Phys. Rev. Lett.* 30, 1343; 1973 and H. D. Politzer *Phys. Rev. Lett.* 30, 1346; 1973). It had been worked out by then that quarks have a quantized property called 'colour charge', which is analogous to electrical charge but comes in three 'colours': red, blue and green. The theory that attempted to explain the colour interaction was called quantum chromodynamics (QCD).

Before these two papers, no one knew how to solve the equations of QCD theory. The problem was the sheer complexity of the interaction. Just as electrical forces between particles are mediated by photons passing between them, so the strong force is mediated by particles called gluons. But whereas photons have no electrical charge, gluons do

have a colour charge, so they interact with each other as well as with quarks. This meant that "the theory was so strongly interacting that it was incalculable", says Forshaw.

Gross, Wilczek and Politzer got round this problem by calculating that the strength of the colour force gets weaker as the particles' energies increase — or equivalently, as the particles get closer together. At sufficiently high energy, the quarks act as though they are 'free'. This became known as 'asymptotic freedom'.

Their work showed particle physicists that if they wanted to understand the strong force they needed to probe it at high energies, where the weakening of the interaction made QCD a testable theory. Studies of high-energy collisions between particles, in particular at the DESY accelerator in Hamburg and the Large Electron-Positron collider at CERN in Geneva, have now verified in great detail the predictions of asymptotic freedom.

Not all aspects of the strong interaction are understood, however. For instance, there is still no way to use QCD to explain some basic properties of quarks at the relatively modest energies characteristic of everyday matter.

But it is precisely the understandable, high-energy region of QCD that researchers are looking at to find a 'theory of everything'. Modifications of QCD at high energies, for example, should be able to describe the Higgs boson, the particle thought to be responsible for giving matter mass.

The Nobel Prize in Chemistry was announced after *Nature* went to press. For coverage see www.nature.com/news

Harvard ceremony graced by hula-hooping laureates

Steve Nadis, Boston

This year's fourteenth Ig Nobel Prize ceremony clashed with the first debate of the US presidential election campaign. But there was no evidence of regret inside the packed auditorium at Harvard University, Massachusetts. For these fans of science that "makes you laugh, and then makes you think", it was George W. Bush and John Kerry who were missing the show.

Presided over by the King and Queen of Swedish Meatballs, as usual, the 30 September event featured Nobel laureates Dudley Herschbach, William Lipscomb and Richard Roberts blowing bullhorns, hula-hooping and singing the classic *You're Just Too Good To Be True* with a karaoke machine.

Herschbach was the delectable prize on offer in the evening's contest to win a date with a Nobel laureate, and Roberts was called on to describe the concept of heredity in just seven words: "Heredity means blame

your parents, not yourself," he said.

The Ig Nobel Prize in Medicine went to Steven Stack of Wayne State University in Detroit, Michigan, and James Gundlach of Auburn University, Alabama, for their paper, "The effect of country music on suicide" (*Social Forces* 71, 211–218; 1992). The physics prize was awarded to Ramesh Balasubramaniam of the University of Ottawa in Ontario, Canada, and Michael Turvey of the University of Connecticut, Storrs, for their sterling work on the dynamics of hula-hooping (*Biol. Cybern.* 90, 176–190; 2004) — it seems that to keep the plastic ring spinning around the waist, it is vital to move up and down at the knees.

The coveted public-health prize went to Jillian Clarke, a Chicago high-school graduate, for experiments that validated the 'five-second rule', which states that it is safe to eat food dropped on the floor if it stays there was sufficiently brief. And the peace

prize went to Daisuke Inoue of Hyogo, Japan, the inventor of karaoke. "Let's party!" he urged the crowd.

Daniel Simons of the University of Illinois at Urbana-Champaign and Christopher Chabris of Harvard University brought home the psychology prize for their report, "Gorillas in our midst", which showed that when observers concentrate on one thing — in their study, people passing a basketball back and forth — they can completely overlook something else — such as a man appearing in a gorilla suit.

And the biology prize went to a collaboration from Canada, Denmark, Scotland and Sweden for demonstrating the role of flatulence in herring communications. Accepting his award, Lawrence Dill of Simon Fraser University, Canada, summed up the evening: "It's been a gas," he said.

▶ www.improb.com/ig

Subtler tests urged for supercomputers

Jim Giles, London

The supercomputer chart will soon have a new number one: data just released by IBM make its Blue Gene/L machine the world's fastest computer.

But some supercomputer specialists, while applauding IBM's technical prowess, question the significance of these rankings. They point out that the technique used to compare supercomputer performance is badly out of date. And they worry that the chart, which guarantees widespread publicity for whichever machine hits the top spot, is skewing US priorities in computer science funding.

IBM says that Blue Gene/L, which is currently on company premises in Rochester, Minnesota, prior to shipment to the Lawrence Livermore National Laboratory in California, can perform 36,000 billion calculations a second. The result has already been submitted to the organizers of the TOP500 list of supercomputers, who will publish their next chart in November. Blue Gene/L is expected to run about ten times faster next year, when all its 130,000 processors are installed.

Alan Gara, the machine's chief architect, says that US government officials are likely to be as pleased with the milestone as IBM is. After the Earth Simulator, built by NEC in Japan, hit the top spot in 2002 (see *Nature* 416, 579–580; 2002), Congress began to pay renewed attention to supercomputer funding. Politicians "want to show that the United States is competitive", says Gara.



Gene genie: IBM's prototype can perform up to 36,000 billion calculations per second.

Yet the researchers who create the chart say that its contents should not be taken too seriously. Machines are ranked by the time it takes them to solve a set of linear equations, a test known as the Linpack benchmark. The test is a good measure of the speed of a computer's processors, says Erich Strohmaier, a computer scientist at Lawrence Berkeley

National Laboratory in California, who helps to compile the TOP500 list. But the test is less sensitive when it comes to assessing the speed at which the processors can communicate with each other — a crucial factor in actual performance, he says.

Using more than one benchmark could help, suggests Dale Nielsen, a theoretical physicist at Lawrence Livermore. A test that involves solving another type of equation, known as a fast Fourier transform, would be a useful addition, adds Strohmaier.

But even if the chart provided a perfect measure of hardware power, it still wouldn't tell the full story. Computers are useless without software, and some scientists feel that the focus on building ever faster machines distracts from the real bottleneck for heavy-duty computer users: the shortage of code that can fully exploit supercomputers' power.

Rick Stevens, a computer scientist at Argonne National Laboratory in Illinois, estimates that some \$500 million a year is spent on funding scientific software development in the United States. But this cash is spread around every scientific discipline, he says, and some areas still lack adequate resources.

At present, Stevens says, researchers often find that code they want to use for a particular application is not optimized for the machine to which they have access. He adds that there has been little investment in adapting code to run efficiently on the fastest supercomputers. ■

Tardy earthquake excites California geophysicists

David Cyranoski

For a group of scientists in the little town of Parkfield, California, the Earth finally moved — when an earthquake of magnitude 6 hit last week. After lying in wait for a major rupture in this part of the San Andreas fault for more than 20 years, scientists in the area pounced on their huge array of monitors to examine the data.

The earthquake, which caused little damage and no deaths, will provide a wealth of information for researchers. But it has also thrown up questions about whether major earthquakes come in predictable periods or with predictable characteristics.

Scientists studying Parkfield in the late 1970s hypothesized that an earthquake would hit the area once every 22 years. The current earthquake, striking after a 38-year gap, might make them recalculate. It does seem to add to the evidence that this earthquake zone may be "quasi-periodic", says David Jackson, an earthquake researcher at the University of California, Los Angeles. But he generally

doubts the idea that certain places have earthquakes with specific characteristics.

Previous Parkfield earthquakes have led scientists to expect foreshocks of up to magnitude 5 striking the area about 17 minutes before the main earthquake. Instead they got aftershocks of magnitude 5. The earthquake also moved north to south along the fault, the opposite of what normally occurs in this region. Showing up late, moving backwards and speaking out of order, the delinquent earthquake is posing a few problems. "These are interesting phenomena, but we don't yet have any explanation," says geophysicist Steve Hickman of the US Geological Survey.

Because of the regularity expected of Parkfield's earthquakes, the area has been intensely monitored for decades. Instruments are used to measure details of structural changes deep below the surface and the movement of the fault. "The data will be copious and sweet," says Jackson.

The earthquake did not come at a great

time for researchers working on the San Andreas Fault Observatory at Depth (SAFOD), also in the Parkfield area. This project aims to drill into the heart of an earthquake-generating zone about 5 kilometres away from last week's rupture to study the fault itself during an earthquake. The project reached a drill depth of 3 kilometres two weeks ago, but seismometers, which would have given unique data, have not yet been installed in the hole. And the project will not drill through the fault itself until next year.

A second, pilot hole a few metres away did have some seismometers in place. But this summer, workers accidentally snipped off 25 seismic stations from the bottom of a string of 32. They have not yet been replaced.

Last week's earthquake has got SAFOD project leaders excited, however. Their part of the fault line is expected to produce smaller earthquakes every two to three years, so they have plenty of time to see more earthquake activity. ■

Blair to seek consensus on safe greenhouse-gas levels

Jim Giles, London

Tony Blair, the British prime minister, plans to adopt a controversial new approach to international negotiations on climate change, according to UK scientists.

The approach, which his government is expected to announce later this month, would ask world leaders to seek agreement on an acceptable target level for the concentration of greenhouse gases in the atmosphere. But climate-change experts in Britain have expressed concern that such a strategy could dilute existing attempts to cut emissions of greenhouse gases, through the implementation of the Kyoto Protocol (see page 613).

With Blair hosting a meeting of the Group of Eight industrialized nations (G8) at Gleneagles in Scotland next July, and Britain holding the rotating presidency of the European Union for six months after that, the prime minister wants to provide some global impetus towards action on climate change.

Blair's initiative would get government leaders to work out how they could declare a level at which atmospheric greenhouse-gas concentrations would become "dangerous", say researchers who have discussed the idea with UK government officials. Supporters of the idea believe that discussion of such a long-term limit could help break the deadlock between countries that have ratified the Kyoto Protocol and others, led by the United States, that have rejected it.

UK scientists who have been consulted on the plan welcome Blair's focus on climate change, but warn that it will be hard to reach political or scientific agreement over what constitutes a dangerous level of greenhouse gases. They say that they have expressed their concerns during consultations for a government-run conference on greenhouse-gas stabilization that will take place next February at the Hadley Centre for Climate Prediction and Research in Exeter.

"Any attempt to launch negotiations on a target would be extremely dangerous and misguided," says Michael Grubb, a specialist in climate-change policy at Imperial College London. He says he received a "stony reception" from environment ministry officials when he made this point to them last month at a meeting at the Tyndall Centre for Climate Change Research in Norwich.

Grubb and others say that consideration of such a greenhouse-gas target will lead different countries in divergent directions, with each one looking at local problems, such as the effect of climate change on staple crops or on the frequency of summer heat-



Tony Blair wants action on global warming.

waves. David Griggs, director of the Hadley Centre — who has also been consulted by the environment ministry — says he fears that nations will find little common ground for identifying such a target.

And Mike Hulme, director of the Tyndall Centre, who is due to meet with ministry officials next week, says that the idea of a global target for greenhouse gases is too distant from people's immediate concerns about the impact of climate change. "It's too remote" he says. "It's not good at changing people's behaviour."

David Warrilow, a ministry official who is involved in the consultation, says that the British government wants other nations to "start to think about progress" towards setting a stabilization level, rather than establishing a firm target. Such debate could feed into future Kyoto Protocol negotiations, he adds.

But some climate-change analysts say the issue will simply divert attention from the pressing need to control emissions of greenhouse gases. "We could invest an enormous amount of time in a fruitless exercise," says Elliot Diringer, director of international strategy at the Pew Center on Global Climate Change in Washington, "rather than focussing on what can be done now."

Diringer agrees that the scope of the Kyoto Protocol should be broadened during its second phase, but he suggests that any new targets should focus on variables over which nations have direct control, such as levels of energy efficiency or the capture and storage of carbon dioxide.

Britain warms to European space exploration plan

Jim Giles, London

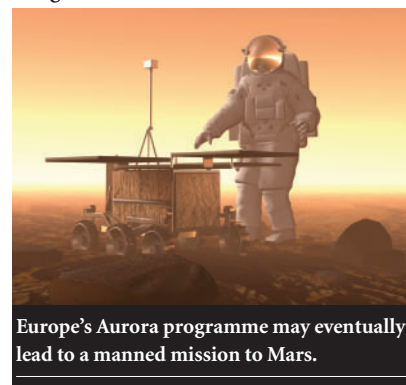
Italy and Britain look set to take a leading role in a major European Space Agency (ESA) initiative to explore the Solar System, which could culminate in a manned mission to Mars.

The involvement of Britain, which pledged €7 million (US\$8.6 million) of initial funding on 1 October, marks a significant re-engagement in ESA projects. Britain did not participate in the agency's work on the International Space Station (ISS), but has recently been taking a more positive approach to ESA. In the past, it has taken a dim view of some of the space agency's more ambitious projects, including its astronaut programme.

Britain will now play a prominent role in the Aurora exploration programme, although the contribution from Italy — €14 million, to be confirmed by mid-October — will give that country the largest single influence on the project. Germany has yet to commit to the programme, and France says it will only supplement the limited funding it has already provided if uncertainties surrounding the future of the ISS can be resolved (see *Nature* 423, 103; 2003).

Details of the programme are due to be agreed by early 2006, but all parties admit that much has yet to be decided. Britain would like Aurora to focus on robotic missions, whereas Italy and France would like it to have a substantial manned component.

Whatever form the plans finally take, the programme will be costly. Britain's Particle Physics and Astronomy Research Council (PPARC), the funding body that pays the UK subscription to Aurora, says ESA may ask it for around €35 million a year. That is probably "wishful thinking" on the space agency's part, says PPARC chief executive Ian Halliday, unless his council receives extra funding from the UK government.



Europe's Aurora programme may eventually lead to a manned mission to Mars.

Space scientists support satellites as aid for Afghanistan

Paris After more than three decades of war and economic strife, the Afghan people can use all the help they can get. Satellite-based education, telemedicine and remote sensing would be a cost-effective way for Western nations to provide support, according to a study from the International Academy of Astronautics.

For an estimated \$750,000, the United Nations or other organizations could set up pilot projects such as two-way video connections between doctors in Kabul and rural health centres. There is currently only one doctor for every 50,000 people in Afghanistan, and the medical charity Médecins Sans Frontières recently pulled out of the country owing to safety concerns.

For about the same amount of money, Afghanistan could use satellites to beam lessons to rural schools, say study leaders François Becker, dean emeritus of the International Space University in Strasbourg, France, and Krishnaswamy Kasturirangan, a member of India's parliament and former head of that country's space agency. Images from remote-sensing satellites could also prove invaluable for planning the new roads and other infrastructure that Afghanistan needs, they add.

► www.iaa.net.org/p_papers/spaceforpeace.pdf

Remote Indian doctors give healthy boost to Pakistan

New Delhi As scientists advocate the use of satellites to help developing nations such as Afghanistan, a similar project has been established in India and Pakistan.

On 28 September, a telemedicine link was established between a state-run medical imaging centre in Lahore, Pakistan, and the private Indraprastha Apollo Hospital

in New Delhi, India. The link will give doctors in Pakistan access to expert medical advice in real time, and the Indian hospital has promised to provide the first 100 teleconsultations free of charge.



Flood of objections: environmentalists want to drain Yosemite's Hetch Hetchy reservoir.

Yosemite dam targeted in valley drainage plan

San Diego An environmental report has added fresh impetus to a campaign to drain a reservoir in Yosemite National Park, California, restoring a canyon flooded by water 90 metres deep to its former glory.

The non-profit campaign group Environmental Defense, based in New York, released a report on 27 September showing how the water in the 80-year-old Hetch Hetchy reservoir could be stored elsewhere without detrimentally affecting San Francisco's water supply. Three other reservoirs already in the Tuolumne River could take up the slack, they say, if supplemented with some new pipes connecting them to the city.

The cost of the programme is estimated at \$500 million to \$1.6 billion. Supporters suggest that this could be factored into the current \$3.6-billion plan to rework the Tuolumne River system. Officials have made no comment about the feasibility of the plan, but say the security of the city's water supply will be their top priority.

Tree-ring study heats up climate-change debate

Munich Temperatures on Earth might have fluctuated more widely during the past 1,000 years than had been thought, according to a recent study. If so, the recent trend of global warming — particularly before 1980 — is less unusual than was previously thought.

The study questions researchers' use of tree rings and ice cores to construct a picture of past global temperatures. Hans von Storch, a climate modeller at the GKSS Research Centre in Geesthacht, Germany, and his colleagues tested this process by creating virtual tree-ring records from a computer model of Earth's past climate. They then used these records to reconstruct global climate. Surprisingly, the temperature fluctuations in the reconstruction were much smaller than those in the original model (A. von Storch *et al.* *Science* doi:10.1126/science.1096109; 2004). Von Storch says the result indicates that researchers might have underestimated the size of past temperature fluctuations by a factor of two or more.

"This is the first critical look I have seen into possible limitations of this widely used methodology," says Chris Forest, a climate researcher at the Massachusetts Institute of Technology. Von Storch's finding doesn't alter the fact that human emissions of carbon dioxide are warming the planet, he says. But, he adds, if the analysis stands up, it will challenge our understanding of how natural climate swings have affected temperature changes.

SpaceShipOne hits the heights to win X prize

San Diego The privately built SpaceShipOne scooped the US\$10-million Ansari X prize for manned, commercial space flight on Monday when it punched into suborbital space for the second time in five days.

The craft, designed by Burt Rutan and his team at Scaled Composites and backed by Microsoft co-founder Paul Allen, reached an altitude of more than 100 kilometres over the Mojave Desert, after being released from a booster plane at about 14 kilometres.

Supporters of the prize hope the flight will pave the way for an era of affordable space travel. But many experts question whether space tourism can really take off before companies have overcome the much tougher hurdle of achieving orbit.

► www.nature.com/news/specials/xprize

Program crosses web to fill in puzzling words

Munich Italian scientists have developed a computer program that they claim can solve crossword puzzles in any language by using the Internet search engine Google to find answers on the web.

Marco Gori and his colleague Marco Ernandes of the University of Siena say the program should be ready for use by the research community by the end of the year. "The idea is not to spoil the enjoyment of players," says Gori — this quixotic use of the program has simply helped them to explore an area of artificial intelligence.

Gori thinks the program will be useful to those who are trying to develop intelligent ways of searching for information online or of scheduling complex interlinked events, for example.



Mentors award

Nature and the UK National Endowment for Science, Technology and the Arts are launching a new award for creative mentorship in science. See page 28 in this week's *Naturejobs* and www.nature.com/nature/nestaaward

Rocket man

One of Saddam Hussein's senior weapons designers tells Geoff Brumfiel of his past, and of his hopes for a new career.

The former Iraqi rocket scientist is eager to show off his photographs. Deftly extracting them from a dusty briefcase beneath his chair, he flips through images of friends, colleagues, and himself standing in the desert in a blue Iraqi Air Force uniform.

But it is with paternal pride that he leafs through pictures of his rockets. He has dozens of photos showing missiles he helped to design shooting skyward, and he can rattle off technical details for each model — how far it flew and what its fuel system was. “We built these out of nothing,” he reminisces.

For more than 40 years, 24 of them under the ultimate supervision of Saddam Hussein, Nadeem (not his real name) was a senior scientist in Iraq's military rocket programme.

On 18 September, he was one of six former Iraqi weapons scientists who flew to Washington for a week of meetings organized by the Civilian Research and Development Foundation (CRDF), a small non-profit organization devoted to keeping former weaponeers from Russia, Iraq and elsewhere in gainful employment by linking them with legitimate private industry in the West.

At the CRDF's headquarters in Arlington, Virginia, Nadeem agreed to talk to *Nature* about his career under Saddam Hussein and his fears for the future. A Sunni Muslim from a well-off family in northern Iraq, Nadeem joined the military straight from high school in 1963. The army sent him to a university in the United States to study mechanical engineering, and he returned six years later to begin work as a rocket designer.

This was a turbulent time in Iraq's history. Following the execution of King Faisal II in 1958, violent coup attempts were frequent. But Nadeem says that weapons researchers were doted on, whoever was in power. “The government looked at us as angels from heaven,” he says. “Computers were not well-known back then, but we had computers.” Researchers from allied countries such as Egypt were brought in to provide additional training in specialist fields such as aerodynamics and thermodynamics.

After Saddam Hussein came to power in 1979, military scientists continued to receive generous funding, Nadeem recalls, but they fell under tighter scrutiny than ever before.



Friends in high places: US troops found a tribute to Iraq's rockets in one of Saddam Hussein's palaces.

“Saddam took care of the scientists, but if you spoke out against him he would execute you,” he says. The headquarters of Iraq's state-owned Military Industrialization Corporation were bugged, Nadeem says. He recalls how, in 1991, security officials overheard a fellow designer criticizing Hussein's invasion of Kuwait. Within days the researcher disappeared, never to be seen again.

In the firing line

Nadeem had risen in this period to a very senior position in the Iraqi rocket programme, but his professional life got more difficult after the 1991 Gulf War. In order to comply with UN sanctions following the war, he says, Iraq shut down the rocket programme and sent him and his colleagues home. He lived off savings at first, and then by buying and selling used car parts, which were in desperately short supply under the UN embargo.

The UN then played an unintended role in reviving Nadeem's career, he says. “The UN brought me back,” he remembers with a laugh. During a 1997 interview with UN weapons inspectors, a senior Iraqi official present was so impressed by Nadeem's knowledge that he rehired him as a consultant to one of Iraq's remaining rocket programmes.

For five years, Nadeem worked feverishly on the design of short-range surface-to-surface missiles, using whatever spare parts he could obtain and often building components from scratch. “We scrounged for what we could,” he recalls. Despite the embargoes, he says, he developed a rocket that could fly upwards of 150 km — the limit permitted in Iraq at the time by the UN. Success won Nadeem and his colleagues a personal

audience with Saddam in 2002 — but colleagues advised him not to go, on account of his outspokenness. “They were afraid I would get them into trouble,” he says.

When the US-led coalition invaded Iraq in March 2003, some of Nadeem's missiles were fired at US forces. As US tanks swept north towards Baghdad, he thought the factory that made the rockets would be among their early ports of call. “We were expecting to be hit by a cruise missile,” he recalls. But no missile came. Instead, looters ransacked the factory's design office. “The Americans could have given the order for a tank to stop at every agency, every university,” he says angrily. “But they did nothing.”

Today, Nadeem works in a mid-level administrative position at the newly formed Ministry of Science and Technology in Baghdad. He isn't happy with the status or the salary of his current position — he earns just half of his former salary — and he has dismal pension prospects, despite four decades in the armed forces. Many of his former colleagues are out of work and some ended up in prison.

Unlike other scientists at the CRDF conference, whose skills in biology or chemistry could be transferred to the private sector, the skills of the rocket-builder have few non-military applications. One CRDF official describes him as “a special case”.

Nadeem says he is uncertain about what will come next. Perhaps he will seek new work outside weaponry. “I'd love to do that,” he says. “But after a 41-year career, it would be hard to do. I would like to work in my field, maybe for NASA — if they need me, I am ready.”

Geoff Brumfiel is *Nature's* Washington physical sciences correspondent.

Chipping in

The production of silicon chips is big business, but with success have come environmental concerns. Geoff Brumfiel meets the people helping the industry to clean up its act.

On the face of it, there is nothing cleaner than a semiconductor manufacturing plant. Dust inside is kept to a few particles per cubic metre to prevent it from clogging up the microcircuits as they are assembled. Workers wear bunny-suits to keep hair and flakes of skin from polluting the environment, and water used in manufacturing is filtered so that it is far cleaner than the water most of us drink.

But even though the working environment is pristine, it's still not exactly eco-friendly. "The reality of a fabrication plant is that it's a huge chemical factory," says Gerald Marcyk, a retired director for research at chip manufacturer Intel. Highly toxic elements, such as arsenic and phosphorus, are used to alter the electrical properties of semiconducting materials. Volatile organic solvents are sprayed onto silicon wafers to remove waste when they are etched with circuit patterns. And the operation of a single plant consumes thousands of megawatt hours in electricity and generates millions of gallons of waste water each day.

Most environmental groups agree that semiconductor fabrication is not the dirtiest manufacturing industry, but it is one of the fastest growing — it has grown by more than 10% a year over the past few decades and currently consists of some 900 plants and hundreds of thousands of workers worldwide. Such rapid expansion has led environmentalists and the US government to turn their attention to the impact the industry is having on the environment and on the health of its employees. But in a sector that is capital- and research-intensive, changes to any working practices to make them more eco-friendly are unlikely to happen overnight.

"I think the industry is a lot more receptive than it was 20 years ago," says Ted Smith, director of the Silicon Valley Toxics Coalition,



The production of silicon wafers for chips consumes huge amounts of water and electricity.

a non-profit environmental group in San Jose, California. But, he adds, more still needs to be done. According to Smith, the environmental track-record of the chip manufacturers has been less than stellar in the past. In California, where the industry originated, one in five toxic-waste sites on the US Environmental Protection Agency's list of high-priority sites waiting to be cleaned up is a current or former semiconductor plant.

Health check

Worker health has also been a problem for the manufacturers. In the 1980s, solvents called glycol ethers were linked to abnormal pregnancies among female employees (M. B. Schenker *et al. Am. J. Ind. Med.* **28**, 639–659; 1995), and since 1996 hundreds of lawsuits have been filed by former employees of IBM, who claim that working in the company's facilities led them to develop cancer. IBM disputes the link with cancer, but in August the allegations led the Semiconductor Industry Association (SIA), a trade group of 95 manufacturers, to announce the largest ever cancer study covering the health records of some 200,000 former industry employees (see *Nature* **431**, 7; 2004).

Many of the problems faced by the industry are legacies of earlier technologies. But as new generations of faster computer chips are designed, companies often have to alter their manufacturing techniques, equipment and chemicals. Such changes can introduce new, and potentially unforeseen, environmental problems, but they also offer a chance to improve individual processes.

These are the challenges and opportunities that motivate Farhang Shadman, a chemical engineer at the University of Arizona in Tucson. Since 1996, Shadman has directed the Engineering Research Center for Environmentally Benign Semiconductor Manufacturing. Sponsored jointly by the National Science Foundation (NSF) and industrial partners, the centre brings together engineers, chemists and physicists from universities around the country who are working on ideas that could someday help clean up the business of making microchips. Although it receives just under \$4 million a year in funding, a fraction of the \$14 billion that SIA members spend on research and development annually, the centre has successfully transferred several technologies to manufacturers.

This year the centre's NSF funding will



Scrubbed up: the plants may be spotless, but how environmentally friendly is the microchip industry?

expire, but Shadman is confident that industry will pick up the slack. “We will have a smooth transition from NSF funding to industry funding,” he says. Reed Content, a senior environmental, safety and health manager at Advanced Micro Devices (AMD) in Sunnyvale, California, and chair of the centre’s industrial advisory board, says that he is working hard to ensure it continues to receive funding next year. “I think the industry would like to see it go forward,” he says.

Improving the environmental performance of a high-pace, high-tech industry requires a special way of thinking, Shadman says. Semiconductor manufacturing is a cut-throat business that over the past few years has been operating on increasingly thin margins. “Unless what we’re proposing will improve performance and reduce cost, we’re not going to get the industry’s attention,” he says.

That often means keeping the solutions as simple as possible. Shadman’s own research group has been working on ways to cut water consumption at the plants. Keeping microchips ultra-clean during the manufacturing process means continually having to wash them free of chemicals.

A clean sweep

To figure out the best ways to save water, Shadman’s team looked closely at how it is used. The researchers found that recycling waste water, as some have suggested, is not always the best idea because purifying it requires a lot of power. “Recycling helps the environment in one way but hurts it in another,” Shadman explains.

Instead, the group focused on making water use more efficient by developing sensors that measure microchip cleanliness, and consulting on new plant designs. The team

has reduced water used for rinsing in newer plants by 20–65%.

More radical changes to the manufacturing process must yield more spectacular benefits if they are to be taken seriously by industry. One such example, according to Christopher Ober, a materials scientist at Cornell University in Ithaca, New York, is the idea of replacing water with ‘supercritical’ carbon dioxide — a hybrid phase of CO₂ that has the properties of both a gas and a liquid.

Supercritical CO₂ forms at temperatures higher than 30 °C and at pressures of several millions of pascals. It flows like a liquid, but has negligible surface tension, giving it some serious advantages over water, Ober explains. For one thing, it can penetrate deep into microcircuitry and dissolve solvents and etching chemicals. Once it has passed through the chip and into a waste system, it reverts to being a gas, depositing whatever waste materials it dissolved. Unlike water, Ober says, this means that it can be easily recycled, and instead of creating gallons of liquid waste it leaves behind a small amount of dry waste. In addition, CO₂ is a by-product of other processes inside the plant, and so is readily available.

But supercritical CO₂ has its disadvantages as well. “The downside is that it has very complicated solubility characteristics, so not everything will dissolve in it,” Ober says. This means that additional chemicals sometimes have to be added. Another problem is that plants need new tools to accommodate the process. As a result, companies have so far been reluctant to replace conventional water with supercritical CO₂. “Even though it is high-tech, the industry is still fairly conservative in terms of making changes,” Ober says.

All that may now be changing for reasons that have nothing to do with the environment.

As the features etched on chips reach the nanometre-scale, they become increasingly difficult to clean with water, Ober says. At such scales, the force exerted by the water’s surface tension is enormous and has the potential to damage tiny circuit patterns. Because supercritical CO₂ has little surface tension, companies are starting to incorporate it into their new plant designs.

In this case, the rapidly changing technology of the manufacturing process is actually encouraging companies to be more environmentally responsible. But things rarely work out that way, so other groups with Shadman’s centre are looking to influence decisions about the next generation of materials needed to make chips even smaller and faster.

“Traditionally, the microelectronics industry has worried mostly about performance,” says Krishna Saraswat, an electrical engineer at Stanford University in California. “What we’re now looking for are materials that give the highest performance and fewest environmental problems.”

That shrinking feeling

As transistors grow ever smaller, Saraswat explains, new materials are needed to make sure that electrons do not leak from one part of the device to another. Several new compounds are being considered, but few people are worrying about their environmental consequences. Saraswat’s work suggests that one class of materials — those involving the relatively benign element hafnium — are the least environmentally offensive.

But if hafnium-related materials turn out to be less effective than the others, their environmental benefits won’t make them the industry’s first choice. Marcyk says that the bottom line will always come first. “The reality is that people want to have an environmentally clean process, but they will not pay any more money for it,” he says. “There are very few changes that I’ve seen that have been done strictly for the environment.”

So can initiatives such as Shadman’s centre really alter the course of semiconductor manufacturing? Some sceptics are less than sanguine about its chances. Smith calls the scale of the research efforts “painfully inadequate”. In an industry where manufacturing techniques are constantly revised, he says, “you’re always going to be playing catch up and clean up.”

Supporters, on the other hand, remain upbeat about the industry’s progress. Technologies suggested by the centre are making their way into plants, and young engineers entering the industry are willing to think about environmental issues. In the end, Shadman argues, the financial pressures that drive the industry will someday force all companies to think about their environmental impact. “Environmental concerns are not a luxury,” he says. “It’s a question of sustainability.” ■

Geoff Brumfiel is *Nature’s* Washington physical sciences correspondent.



Starting from scratch

Genetic engineering is old hat. Biologists are now synthesizing genomes, altering the genetic code and contemplating new life forms. Is it time to think about the risks? Philip Ball asks the experts.

Redesigning Life. That was what Steven Benner wanted to call his 1988 conference in Interlaken, Switzerland. A chemist now at the University of Florida in Gainesville, Benner was organizing the meeting to explore the possibilities for making artificial chemical systems that mimic essential features of living things.

But his title caused such a furore among prospective attendees that Benner had to tone it down to *Redesigning the Molecules of Life*. “Individuals as distinguished as Nobel laureates were convinced that the title would incite anti-recombinant-DNA riots in Switzerland,” Benner explains.

Benner’s conference helped to define one strand of the emerging discipline known as synthetic biology, a field that is now raising worries that won’t be deflected simply by semantics. The expanding toolbox of ways to re-engineer microbes — and even construct new ones — has opened up extraordinary possibilities for biomedical discovery and environmental engineering. But it also carries potential dangers that could eclipse the concerns already raised about genetic engineering and nanotechnology. If biologists are indeed on the threshold of synthe-

sizing new life forms, the scope for abuse or inadvertent disaster could be huge.

In a dramatic demonstration of the potential risks, virologist Eckard Wimmer at the State University of New York at Stony Brook announced in 2002 that his team had built live poliovirus from scratch using mail-order segments of DNA and a viral genome map that is freely available on the Internet¹. The feat put a spotlight on the possibility that bioterrorists could create even more dangerous organisms — including Ebola, smallpox and anthrax — perhaps endowing them with resistance to antibiotics.

Creative thoughts

Since then, biologists’ abilities to engineer life have bounded ahead. Wimmer took three years to build his poliovirus, but last November genome sequencer Craig Venter and his colleagues at the Institute for Biological Energy Alternatives in Rockville, Maryland, announced that they had taken just three weeks to assemble a virus that infects bacteria². At the same time, bacterial cells are being rewired to perform functions they can’t fulfil in nature. And researchers are getting close to determining the smallest

set of genes necessary to support a living cell, which might make it possible to cook up new life forms.

Almost 30 years ago, concerns that recombinant DNA technology could pose risks to human health and the environment prompted leading molecular biologists to call an unprecedented summit. They gathered at the Asilomar Conference Center in Pacific Grove, California, in February 1975, where they decided to voluntarily forego some kinds of research and to instigate safety measures to prevent abuses of the new techniques.

Is it now time for another Asilomar? Researchers involved in synthetic biology generally agree that more discussion of how to avoid risks is urgently needed, but have yet to take the formal step of calling for a summit. Some concerns were aired at a special session at the First International Meeting on Synthetic Biology, held in June at the Massachusetts Institute of Technology (MIT) in Cambridge, but it did not set out to produce policy recommendations.

The reason we face the question of risk at all is that the potential rewards of pursuing synthetic biology are so great. Protein engineer Wendell Lim of the University of

California, San Francisco, says that if synthetic biology is successful, it may become possible to treat a variety of diseases by repairing defective cell functions, targeting tumours or stimulating growth and regeneration of specific cell types. Other researchers are hoping to engineer bacteria to make complicated drugs or to use sunlight to generate clean-burning hydrogen for cars and power plants.

Synthetic biology is the logical corollary of the realization that cells, like mechanical or electronic devices, are exquisitely 'designed' — albeit by evolution rather than on the drawing board. Their functions are enacted by circuits of interacting genes. As scientists began to map these circuits in the 1990s, they inevitably began to wonder whether they could rewire them.

Glowing report

In 2000, biological physicists Michael Elowitz and Stanislas Leibler, both then working at Princeton University in New Jersey, designed from scratch a genetic circuit that caused oscillating production of a fluorescent protein. Bacteria programmed with the circuit glowed periodically³. Other researchers built on this, creating circuits that could be switched on and off by external signals, or that could control bacterial population density^{4,5}.

Now a growing number of researchers are working on ways to alter the circuitry of cells. Lim, for instance, is retooling some of the proteins that carry signals within and between cells so that they respond to different inputs from the environment^{6,7}. And chemical engineer Jay Keasling at the Lawrence Berkeley National Laboratory, has refitted the gut bacterium *Escherichia coli* with the circuitry it needs to synthesize a precursor to the powerful antimalarial drug artemisinin, a product of the wormwood plant that is currently too expensive for widespread use. This meant importing ten genes from other organisms, including wormwood and brewer's yeast, and then carefully tuning their expression levels⁸. If this proves to be a cheap, reliable source of the drug, it could transform the treatment of malaria.

In a parallel development, other researchers have been tinkering with the building blocks of genes and proteins themselves. Naturally occurring proteins are built from a standard set of 20 amino acids. Although these are enough to produce protein chains with a staggering array of functions, expanding this repertoire might enable the design of biomolecules with new functions, such as protein-based drugs that resist being broken down in cells.

In 1989, Peter Schultz, a chemist now at the Scripps Research Institute in La Jolla, California, reported that he had found a way to persuade bacteria to incorporate



Tooled up: Steve Benner (above) and Wendell Lim are working to redesign proteins and the DNA in living cells.



an unnatural amino acid into a specific protein⁹. This produced enzymes with subtly different activities. Since then, Schultz has added more than 80 unconventional amino acids to proteins.

Culture shock

In the same year, Benner persuaded cells to insert a base pair not used in nature into their DNA¹⁰. A better understanding of the different types of molecules that can function as DNA bases will open a window to the possible chemical ancestors of DNA that might have existed on primordial Earth, and to the possible genetic systems that could support life on other worlds. "I suspect that, in five years or so, the artificial genetic systems that we have developed will be supporting an artificial life form that can reproduce, evolve, learn and respond to environmental change," Benner predicts. "This will help define how life not of earthly origin might appear."

As biologists learn to shape cellular circuits and their molecular components, developments in the automated chemical synthesis of DNA are allowing entire genomes to be designed and assembled. Venter's lightning-fast synthesis of a virus in November was a testament to the expanding capacity of DNA synthesis machines. By some estimates, next year's machines will be able to generate sequences about a million base pairs long — roughly the size of the genome of *Chlamydia*, which causes a common sexually transmitted disease, and a quarter the size of *E. coli*'s genome.

"Bacterial genomes are within the range of current DNA-synthesis technology," says John Mulligan, president of the DNA-synthesizing company Blue Heron Technology in Bothell, Washington. But bacterial genomes must be embedded within a cell and its attendant biochemical machinery, making them much harder to synthesize than viruses. Nevertheless, attempts are under way. In November 2002, Venter made a high-profile announcement of his intention to build a simple bacterium starting with machine-made DNA.

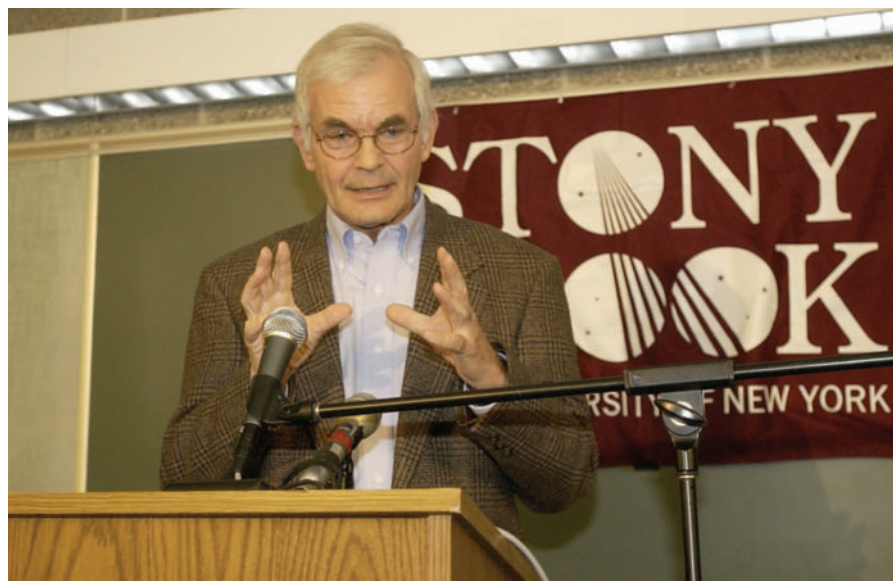
Plain and simple

But building a new bacterial genome is not just a matter of chemistry — you have to design the circuitry too. That's the hard part, so it's good to simplify. "An alternative to understanding complexity is to get rid of it," says Tom Knight, a computer scientist at MIT who brings an engineer's perspective to synthetic biology.

To this end, Knight is studying one of the simplest organisms known, *Mesoplasma florum*, a bacterium that has only 682 genes. The draft genome of this organism was completed last year, and its metabolic pathway has been mapped. The 793-kilobase genome seems to contain very little non-essential DNA, but Knight thinks it can be simplified further. He is now mapping its circuitry and modelling it on a computer to see what else can be removed.

All of these technologies combined are raising issues similar to those that sparked the Asilomar summit. Back then, molecular biologists realized they had all the tools to genetically modify bacteria — and possibly higher organisms — in just about any way imaginable. The hope was that bacteria could be engineered to produce drugs such as human insulin cheaply, and indeed they soon were. The worry was that no one knew how modified bacteria might fare in the environment — whether, for example, they might be toxic, or resistant to antibiotics.

Synthetic biology is now raising the bar. Should limits be set on what is attempted? If so, what should they be and how should they be enforced? And what steps can be taken to ensure that a rogue organization, or even a state-sponsored bioweapons



Past the post: Eckard Wimmer announces the complete synthesis of poliovirus from mail-order DNA.

programme, does not use the technology to synthesize a dangerous microbe?

Roger Brent, president of the Molecular Sciences Institute in Berkeley, California, suggests that one option might be for DNA synthesis to require a licence. But more importantly, Brent says, synthetic biology should avoid developing a hacker subculture like that which spawns computer viruses. Rogue computer hackers hope to earn respect from their peers by producing particularly clever or insidious virus programs. Brent urges researchers in the field to encourage responsible lab culture by not engaging in showy stunts with no research purpose.

Assembly lines

Even though licensing is currently not required, some DNA synthesis companies have taken their own steps to avoid inadvertently aiding irresponsible work. Molly Hoult, senior vice-president of Blue Heron, says that all the company's orders for DNA are cross-checked against a database of "biological nasties". If a match turns up, the company tries to find out more about the customer's research before completing the order. If it can't easily be checked out, Blue Heron simply turns the order down. "We walk away from some business," Hoult says.

Such self-policing could become the norm, and scientists might even be asked to cooperate more closely with intelligence agencies to prevent the abuse of synthetic biology. An unclassified report by the CIA released last November warned that synthetic biology could produce engineered agents "worse than any disease known to man" and suggested "a qualitatively different working relationship between the intelligence and biological sciences communities". In particular, the bioscience community might function as a "living sensor web" that reports to the government on technical



The CIA is calling for a change in the way the intelligence service and scientists work together.

advances that could be used as weapons¹¹.

But it is not clear whether the risk of bioterrorism will be the most important concern with synthetic biology. Ron Weiss, an electrical engineer at Princeton University who spends his time rewiring bacteria, points out that adding antibiotic-resistance genes to harmful bacteria is relatively straightforward and has been possible in principle since the 1970s — yet it has not become a major focus of biowarfare. It would be easier and cheaper simply to breed and release existing harmful organisms than to make new ones. "If I was a terrorist," says Weiss, "this isn't the way I'd get maximal damage for my buck."

It is much harder to anticipate the unintentional dangers of synthetic biology. For example, if new strains of bacteria were developed with unprecedented capabilities, how could they be kept under control?

One way might be to use built-in safeguards. For instance, the innate ability of bacteria to respond to high population density, a feature known as quorum sensing, could be co-opted to activate a self-destruct mechanism. Another option might be to build gene circuits that function like the logic gates of computers to count the number of times a cell divides. After a preset number, the cell would die.

Initial attempts have been made. Unfortunately, Weiss has found that mutant strains evolve after just a few days that can evade his population-control mechanism⁵. But he thinks this can be solved by creating several layers of defence. After all, such redundancy seems to be built into naturally occurring quorum-sensing bacteria, which do not mutate to evade their own population controls. "Nature does this already," Weiss says.

Into the unknown

Yet as synthetic biology develops, it will be hard to anticipate all the possible problems, whether malevolent or inadvertent. "The repertoire over the coming decade is limitless," says George Poste, a bioterrorism expert and director of the Biodesign Institute at Arizona State University in Tempe. "You'll never identify all the risks." Poste says that he is not particularly concerned about immediate dangers, as most researchers are still working with biological materials isolated from cells, so nothing is likely to escape from the laboratory. But "fast-forward two decades and it may be quite different", he adds.

To help quantify risks as they emerge, Poste proposes developing what he calls a 'calculus of risk' — an equation that can enumerate a 'risk factor' for new developments and sound an alarm bell when a certain risk threshold is reached. It's a necessarily crude tool — Poste's equation includes poorly quantified factors such as the projected time it would take to convert a new technology for malevolent use — but it might at least help to distinguish remote risks from more immediate ones.

The difficulty of putting a finger on the risks might leave researchers attending an Asilomar-style conference clutching at shadows. So for now the talks will remain informal. "This definitely merits a lot more discussion," says Weiss. "We don't understand the issues sufficiently yet."

Sooner or later, synthetic biology may find itself facing dangers that are far more than hypothetical. As Poste puts it: "Biology is poised to lose its innocence."

Philip Ball is a consultant editor of Nature.

1. Cello, J., Paul, A. V. & Wimmer, E. *Science* **297**, 1016–1018 (2002).
2. Smith, H. O., Hutchison, C. A., Pfannkuch, C. & Venter, J. C. *Proc. Natl Acad. Sci. USA* **100**, 15440–15445 (2003).
3. Elowitz, M. B. & Leibler, S. *Nature* **403**, 335–338 (2000).
4. Gardner, T. S., Cantor, C. R. & Collins, J. J. *Nature* **403**, 339–342 (2000).
5. You, L., Cox, R. S., Weiss, R. & Arnold, F. H. *Nature* **428**, 868–871 (2004).
6. Dueber, J. E., Yeh, B. J., Chak, K. & Lim, W. A. *Science* **301**, 1904–1908 (2003).
7. Park, S.-H., Zarrinpar, A. & Lim, W. A. *Science* **299**, 1061–1064 (2003).
8. Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D. & Keasling, J. D. *Nature Biotechnol.* **21**, 796–802 (2003).
9. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. *Science* **244**, 182–188 (1989).
10. Switzer, C., Moroney, S. E. & Benner, S. A. *J. Am. Chem. Soc.* **111**, 8322–8323 (1989).
11. *The Darker Bioweapons Future* (Office of Transnational Issues, CIA, OTI SF 2003-108, 2001); <http://www.fas.org/irp/cia/product/bw1103.pdf>

Brazil needs action rather than words

The downward trend in support for postgraduates could reverse recent improvements.

Sir—As David King reveals in his Feature “The scientific impact of nations” (*Nature* **430**, 311–316; 2004), Brazil compares unfavourably with most other countries. It produces a meagre 1.2 % of the world’s publications, despite having roughly 3% of the world’s population and income. If Brazil is to have any chance of improving its status, the current trend of decreasing financial support for science graduates should be reversed, as a matter of urgency.

During the past few decades, many more people have taken part in Brazil’s graduate programmes. Brazil has a formal system for evaluating these courses, and despite many flaws (as noted by L. de Meis, M. S. do Carmo & C. de Meis, *Nature* **424**, 723; 2003), this has allowed some progress in science and technology.

Such progress has become evident in the increasing number of contributions made to international science by Brazilian scientists — up from 48,800 authors appearing in publications indexed by the ISI in 1998 to 58,000 in 2002. A closer

analysis reveals that this increase was due exclusively to the work of postgraduate authors (who increased from 11,300 to 21,800 during that period), while the participation of career investigators has decreased. Thus, the significance of postgraduate students in the production of scientific knowledge is unquestionable.

However, in a country where science funding is mainly public, the resources allocated by the Brazilian government to science and technology have decreased over this same time period. Between 1998 and 2002 the National Council for Science and Technology Development (CNPq), a federal agency, cut the amount of money allocated for postgraduate fellowships by 50%. This loss of funding is even more drastic when you consider that the number of postgraduate students increased by 36% during the same period. Sadly, this downward trend has also been followed by other Brazilian agencies, such as CAPES (which funds graduate programmes in Brazil and Brazilian postgraduates studying abroad) and FAPESP, which mainly supports

innovation in science and technology.

Inflation has further eroded the value of the remaining fellowships; in 1994, a PhD fellowship provided the equivalent of US\$1,000 a month, whereas today it is worth a mere US\$350.

Given these numbers, it is not far-fetched to say that these students are being exploited. How long can even the most motivated student be expected to survive on just US\$350 a month?

All of these factors are seriously jeopardizing Brazilian science. Surprisingly and contradictorily, the Brazilian government has historically emphasized the importance of science in the development of the country. It is time to convert those words into actions.

André Frazão Helene*, **Verónica S. Valentinuzzi†**

**Departamento de Fisiologia, Instituto de Biociências, Universidade de São Paulo, São Paulo 05508-900, Brasil*

†Departamento de Fisiologia, Centro de Biociências, Universidade Federal do Rio Grande do Norte, Natal 59078-970, Brasil

No mistake in Berkeley’s biotechnology deal

Sir—I still do not understand why the funding deal between the Swiss firm Syngenta and Berkeley’s department of plant and microbial biology was considered “a mistake”, as the Busch report claims (*Nature* **430**, 598; 2004).

Let’s look at the balance sheet. On the up side, Berkeley got \$25 million, which funded 26 researchers. These researchers made 12 patentable discoveries with Syngenta funding, and the company is pursuing six of these. Moreover, there was no harm to the quality or freedom of academic research. On the down side, an individual faculty member may or may not have had problems separating his personal interest from his duties on a tenure committee, and the Berkeley researchers were not especially productive with the Syngenta funding.

Where is the mistake here? It looks to me as if Berkeley got a pretty good deal, and the negatives seem to be the responsibility of individuals in the Berkeley community.

Is there more to the story? Does the report show that Syngenta’s funding somehow caused the disappointing productivity of the research?

Judging from what I have read, it

seems that the deal was a “mistake” only because the critics have decided that all such deals must be mistakes.

Philip G. Hultin

Department of Chemistry, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2

Scientists and teachers should ignore politics

Sir—We read with interest your News story “Nobel laureates spearhead effort to put Kerry in the White House” (*Nature* **430**, 595; 2004) about scientists campaigning in the United States.

Nobel laureates have more productive ways to benefit society than entangling themselves in the chaotic web of political campaigns. The expertise of the American Nobel laureates is in physics, chemistry, physiology or medicine — otherwise they would have received their Nobel Prizes for peace efforts. Why should their expertise in their chosen subjects make them masters in the sphere of politics?

We witnessed an example of this activity in Taiwan’s 2000 presidential elections, and we find reasons for concern. Taiwan’s only Nobel laureate, the chemist Lee Yuan-tseh, supported the Democratic Progressive Party nominee, Chen Shui-bian, before Chen’s election to office in 2000.

After that election, Lee continued to speak out on political matters, with sometimes mixed results. Lee has been involved in educational reforms in Taiwan for more than a decade, but last year university lecturers launched a signature campaign, asking him to take responsibility for what they considered to be failures in educational reform.

Yet Lee’s scientific achievements have been rightly acclaimed. He has also been honoured as a Chinese scholar by the People’s Republic of China. Perhaps his greatest skill is in inspiring a younger generation to become teachers themselves. Lee’s reported words at a recent award ceremony for teachers in Taipei are worth recalling: “It is teachers, not politicians, who control the lifeline of society.”

We recently heard that the 82-year-old Chinese scientist Yang Cheng-ning, who received the 1957 Nobel Prize in physics and has lived for many years in the United States, now teaches physics at Tsinghua University in Beijing as well as continuing his research. This is what laureates should be doing, not taking part in politics.

Minna J. Hsu*, **Govindasamy Agoramoorthy†**

**Department of Biological Sciences, National Sun Yat-sen University, Kaohsiung 40424, Taiwan*

†Department of Applied Foreign Languages, Tajen Institute of Technology, Yanpu, Pingtung 907, Taiwan

The price of hope

What's the benefit to drug companies of making medicines for the poor?

Strong Medicine: Creating Incentives for Pharmaceutical Research on Neglected Diseases

by Michael Kremer & Rachel Glennerster
Princeton University Press: 2004. 152 pp.
\$24.95, £15.95

Pierre Chirac

Access to essential drugs for poor people in developing countries has made headlines several times during the past few years. Most of the media attention has centred on a few specific topics within this broad theme: the debate about access to antiretroviral drugs against AIDS; the hot economic and political topics of patents and the World Trade Organization agreements; and the lack of research and development (R&D) to create desperately needed new drugs for 'neglected diseases', conditions that predominantly affect poor people in developing countries.

The central theme that links these issues together is how to encourage research and development for new drugs and ensure that everyone has quick access to the medicines they need. Only a few years ago the sole paradigm, recognized not only by industry but also within most academic circles, was 'patent + market → R&D'. But now alternative or at least complementary paradigms are being considered. Michael Kremer and Rachel Glennerster, the authors of *Strong Medicine*, favour the latter route. In their approach, they fall between those who still think (or at least claim) that the current system only needs minor adjustments and those who want to see complete change on a global scale.

Kremer and Glennerster's aim is to find practical ways to restart R&D for these neglected diseases. They believe that market and public bodies have failed to motivate R&D — only 1% of drugs launched onto the market during the past 25 years target such diseases. They are convinced that appropriate policies and public commitment could motivate private investment in this area. This strategy has already been followed with rare diseases, through 'orphan drug' legislation passed in rich countries, which grants tax incentives and guaranteed periods of exclusivity to companies that invest in R&D for drugs to treat rare conditions.

The authors comment on the various tools that have been suggested or tested to influence R&D. 'Push' mechanisms refer to financing research inputs such as grants to academics, or tax credits for specific R&D activities. 'Pull' solutions mean financing research outputs, such as a commitment to create a profitable market for the expected



Cost of living: who will pay for new drugs to combat 'neglected diseases' such as sleeping sickness?

products. Kremer and Glennerster tend to favour pull solutions, because these are less bureaucratic and are closer to market-like solutions (which the authors prefer in general). After a short discussion of various pull solutions, the authors opt for one in which somebody creates the market by committing to purchase a certain quantity of drugs at a fixed price.

And this is where this book becomes really original. Kremer and Glennerster enter into a computation for assessing the amount of sales that would be enough to motivate private companies to embark on such a deal with a sponsor. Their conclusion is that the system could work with a guaranteed price of \$15–20 per complete vaccination for the first 150–200 million individuals vaccinated, the rest of the vaccine supply being priced at a modest mark-up over production cost.

Who should pay for all this? Kremer and Glennerster believe that this kind of spending would be popular with governments of rich countries, as public money would only be invested in concrete outcomes — not in research that might lead nowhere. They also mention the World Bank and the Bill and Melinda Gates Foundation (from which they acknowledge financial support). Finally, they believe that a modest co-payment from recipient countries would be desirable.

The book does have some limitations. First, the reasoning is mainly based on vaccines and focuses mostly on the three biggest killers: AIDS, malaria and tuberculosis. These targets are worthwhile, but they are also the easiest, because vaccines generally

need only a few jabs to create a long-term benefit, and because hundreds of millions of people are affected by these diseases. It may be more difficult to convince the public and private sectors to do R&D for the most neglected diseases, such as sleeping sickness or Buruli ulcer.

Another important limitation is that the solution proposed by Kremer and Glennerster is mainly a tool. They recognize that a long-term political commitment will be necessary, but they do not elaborate on how to achieve this. This will frustrate those who believe that such a long-term commitment needs to take the form of an international framework or treaty for neglected diseases. Some people within the World Health Organization and Médecins Sans Frontières remember how difficult it was to mobilize the few million dollars necessary to secure the production of a new combination vaccine against meningitis, just a few months before the 2004 epidemic in Africa. It would have been much easier if some kind of global R&D fund had been available.

Thus, some people might criticize this book for stopping mid-way. But Kremer and Glennerster's solution does imply that the private drug R&D activity could be more appropriately directed by public will and money — something that others might consider a step too far already.

This book should interest anyone involved in international public health, politics and economics. It is a valuable effort to find a practical solution to a major problem, and most readers, when they've finished the

book, will probably say "Let's just try it!" ■
Pierre Chirac is a public-health pharmacist with the Médecins Sans Frontières Access to Essential Medicines Campaign, 2 rue Chaligny, 75012 Paris, France.

The other evolutionist

The Heretic in Darwin's Court: The Life of Alfred Russel Wallace

by Ross Slotten

Columbia University Press: 2004. 648 pp. \$39.50, £26.50

An Elusive Victorian: The Evolution of Alfred Russel Wallace

by Martin Fichman

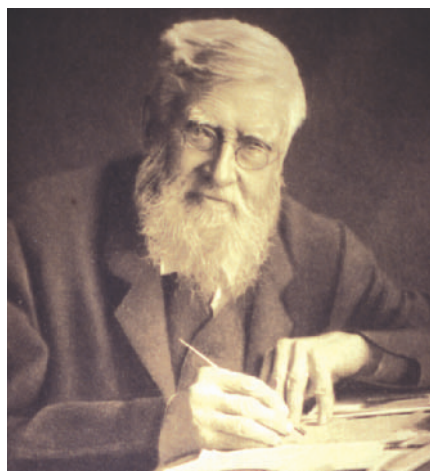
Chicago University Press: 2004. 416 pp. \$40, £28

George Beccaloni

Lying ill with fever on the remote Mollucan island of Halmahera in February 1858, the English naturalist Alfred Russel Wallace puzzled over the phenomenon of 'species transmutation' — a subject that had rarely been absent from his thoughts for the past 13 years. Three years earlier he had published an important but largely unnoticed paper on this topic, now known as evolution, in which he strongly argued that organisms must have evolved from earlier forms. But the mechanism of this process had so far eluded him — as it had (virtually) everyone else.

Suddenly, a flash of insight led him to the idea of natural selection as the mechanism driving evolutionary change. Once he had recovered enough strength, he fleshed out his ideas in an essay entitled "On the Tendency of Varieties to Depart Indefinitely from the Original Type". Knowing that his occasional correspondent Charles Darwin was also keenly interested in the 'species problem', he sent the essay to him on the next mail boat that departed from the neighbouring island of Ternate. Unknown to Wallace, Darwin had in fact discovered natural selection some 20 years earlier and, urged on by the geologist Charles Lyell, he was slowly writing a huge tome in which he planned to present his 'heretical' theory to the world.

The events that unfolded following Darwin's receipt of Wallace's essay have become legendary. Realizing that Wallace had independently reached the same conclusion as himself, Darwin was thrown into a state of confusion and despair. He wrote to Lyell: "Your words have come true with a vengeance ... all my originality will be smashed." Lyell contacted another of Darwin's influential friends, the botanist Joseph Hooker, and together they sought to remedy the awkward situation as best they could.



Alfred Russel Wallace wrote some 700 articles and 20 books on a wide range of subjects.

Without first asking Wallace's permission, they arranged for his essay to be read at a meeting of the Linnean Society in July 1858, along with an abstract of an unpublished manuscript on natural selection that Darwin had written in 1844, and an abstract of a letter outlining the concept of descent with modification that Darwin had sent to the American botanist Asa Gray the year before. By arranging these texts in the order in which they had been written, Lyell and Hooker secured Darwin's priority, and when they were published a month later, Darwin's name naturally appeared as first author of what became the first publication to explicitly propose the theory of evolution by natural selection. This, together with the fact that Darwin's *Origin of Species* (an abstract of his "big book") was printed only a year later, are two of the reasons why Wallace often receives little or no recognition for the discovery.

The Heretic in Darwin's Court by Ross Slotten and *An Elusive Victorian* by Martin Fichman are the latest in a recent spate of books that examine Wallace's fascinating life and often controversial work. Slotten's book is a conventional, chronologically arranged biography, whereas Fichman's aims to provide "a contextualist and analytical study of Wallace's major intellectual and cultural views and activities".

Slotten's book is the most detailed study of its kind published to date and provides a vivid account of Wallace's rich 90-year life. It covers his early impecunious years at school in Hertford (he left school aged 13) and his training as a land surveyor; his four years collecting biological specimens in the Amazon basin (many of which were destroyed when his ship sank on the way back to Britain); his eight years collecting in Southeast Asia (he sent back 126,000 biological specimens, including 1,000 species new to science); and his years in England, during which he wrote some 700 articles and 20 books on an eclectic range of subjects and made a prodigious number of seminal contributions to the fields

of biology, geography, geology and anthropology, among others. Slotten attempts to produce a "three-dimensional portrait of a man whose forays into spiritualism, socialism, antivaccinationism, and other unorthodox 'isms' have been caricatured, overanalyzed, or ignored by specialists in the academic world". However, like most previous biographers, Slotten seems perplexed by some of Wallace's seemingly crackpot beliefs, and although he discusses them at length, he makes little attempt to analyse them in depth.

Fichman, in contrast, tries to do just that and adopts a thematic approach to scrutinize each of Wallace's major interests in turn. He argues that Wallace sought to integrate his beliefs into a single internally consistent world view, which he calls Wallace's "teleological evolutionary cosmology", and he maintains that many of Wallace's putative unorthodoxies are in fact artefacts of historiography. From a modern perspective, it may seem inappropriate for an 'objective' scientist to espouse beliefs such as spiritualism, as Wallace did, but in the late nineteenth century and the early part of the twentieth it was entirely acceptable for scientists to do so.

Fichman's book is the more challenging read and is likely to appeal mainly to serious Wallace scholars. But to understand why a book like Fichman's is needed in the first place, anyone unfamiliar with Wallace's work should probably study a biography such as Slotten's. ■

George Beccaloni is an entomologist at the Natural History Museum, Cromwell Road, South Kensington, London SW7 5BD, UK. He organized a project to restore Wallace's neglected grave.

Cycles of life

Nutrient Cycling and Limitation: Hawai'i as a Model System

by Peter Vitousek

Princeton University Press: 2004. 232 pp. \$79.50, £51.95 (hbk); \$35, £22.95 (pbk)

David Schimel

The regulation of growth by nutrient availability and the control of cycling of the elements through living communities are basic ideas in the field of ecology. Nutrient cycling, or biogeochemistry as it is known, is now a rich research field with a vast literature linked to disciplines as disparate as ecology, soil science and atmospheric science. It forms a bridge between the life and earth sciences, both of which are concerned with the cycling of elements but for vastly different reasons.

The ecological or process-oriented approach to nutrient cycling largely traces back to seminal work in New England in the 1960s and 1970s. In that oeuvre, Peter Vitousek's work for his doctoral thesis began

A world of learning

The small publishing house Eichborn regrets that “whereas the Bible is sold off at any supermarket, Alexander von Humboldt’s counter-theories to the holy writings are hard to find even in the largest of book shops”. In the hope of correcting this, Eichborn is republishing three of the explorer-scientist’s most important works.

Von Humboldt, born in 1769, remains one of Germany’s most important intellectual figures. Deeply committed to Enlightenment principles, he was a collector of knowledge. He travelled extensively, learning the language of each country he visited. He was an internationalist, and persuaded Prussia’s King Friedrich Wilhelm IV to inaugurate, in 1842, the *Order Pour le mérite für Wissenschaften und Künste*, the first such civilian society in war-torn Europe to recognize individual academic achievement across borders.



His massive tome *Kosmos* was an attempt to synthesize all the knowledge of the physical world, from thousands of different sources, including his own scientific travels. It was illustrated with maps like the one shown here, describing not only geography but also geology,

biology and anthropology. *Kosmos* was a thirty-year labour of love, and von Humboldt was reportedly still adding to it on his dying day, aged 89. The work was a best seller in the nineteenth century, and was translated into many languages. But it faded from view in the twentieth century, and few Germans today have read the original words.

The publication of *Kosmos* marks the 200th anniversary of Humboldt’s return from his five-year exploration of the Americas. The two companion volumes are *Ansichten der Natur* (*Views of Nature*) and the sumptuously illustrated *Ansichten der Cordilleren und Monumente der eingeborenen Völker Amerikas* (*Views of the Cordilleras and Monuments of the indigenous American Peoples*), which has, in fact, never previously been published in German. **A.A.**

www.eichborn.de

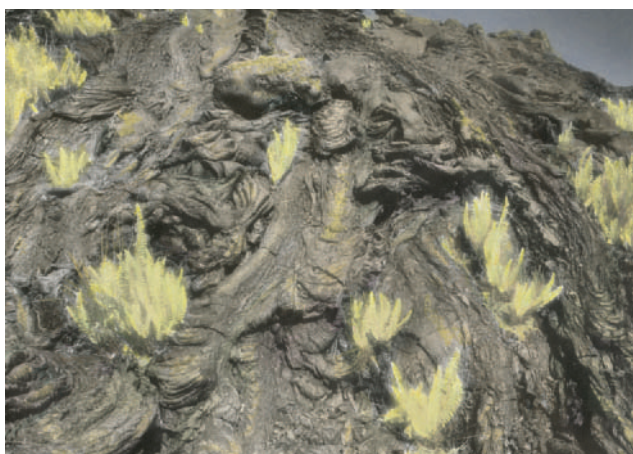
to place nutrient cycling in a mechanistic rather than a budgetary framework. The concept of the limitation of nutrients was central, defined as “occurring when the addition of an essential element increases the growth of individual organisms or populations, or increases the rate of a biological process” (F.S. Chapin *et al.* *Am. Nat.* 127, 48–58; 1986). Nutrient limitation had long been known to foresters and agronomists as a spatially variable property but was usually treated as a static site characteristic. Following the publication of the nutrient-retention hypothesis by Vitousek and William Reiners (*BioScience* 25, 376–381; 1975), nutrient availability was viewed as an emergent property of an ecosystem, a dynamic property controlled by biological, hydrological and geological interactions.

I vividly remember Vitousek stating in a (probably beer-fuelled) conversation about research focus and scientific success that his focus was on “understanding nutrient limitation during succession”. His work began by understanding the interactive controls over nutrient limitation after forest disturbance. In *Nutrient Cycling and Limitation*, Vitousek explores this theme on a grand canvas, basing it on his own work and that of a small army of students and collaborators.

The island of Hawaii is formed from lava flows of diverse and known ages spanning many millions of years. Ecosystems have developed on soils formed from these lava, drawing from a relatively common palette of species but under vastly different conditions of nutrient limitation, depending on the age of the soils. Vitousek reports on several decades of research using age gradients

(along with climatic gradients and manipulation and natural disturbance within the geological template) aimed at understanding how nutrient limitation has evolved as the lava soils of Hawaii have aged over the past 5 million years or so.

He expands the concepts of nutrient cycling, inputs and losses on the ecosystem disturbance and recovery time scale (from 10 to 1,000 years) to geological time. Pro-



Life goes on: plants recycle nutrients to grow from a lava flow.

cesses that on ecosystem time scales are either too slow to matter, or too episodic, come to dominate on geological time scales. Expanding the time scale expands the spatial scale as well: the parsimonious explanation for some aspects of Hawaiian geochemistry turns out to be the transport of nutrient-containing dust across the Pacific, a process that occurs at low instantaneous rates and during climatic episodes separated in geological time, but that is essential to budgets over millions of years. Today’s Hawaiian ecosystem dynamics, though, are controlled partly by nutrient budgets and

ratios reflecting these inputs over the ages.

While *Nutrient Cycling and Limitation* builds a strong bridge to the geological view of biogeochemistry, and to geological time scales, it also extends the biological paradigm. In geochemistry, organisms often appear as essentially passive transducers of physical forces into chemical dynamics. In this view, although rates might differ in the presence or absence of organisms, the geophysical environment is viewed as shaping the system.

Vitousek’s view of nature is richer by far. He shows how the biogeochemical characteristics of organisms can alter the trajectory of geochemical development, on multiple time scales. He views these organismal characteristics as being shaped by the organisms’ deep evolutionary history. When organisms migrate to or invade Hawaii, the conflict between their evolutionary history and that of the island can create new biogeochemical conditions and complex responses. This leads to a contingent rather than an environmentally deterministic view of ecological dynamics, and is

an important counterpart to the simpler views often espoused in biogeography.

This short review does not do justice to the range of topics addressed in *Nutrient Cycling and Limitation*, nor does it convey the depth of theoretical perspective couched in reams of hard-won experimental and observational data. This book will reward reading and re-reading, and is an excellent introduction to biogeochemical ecology for those coming from other fields of science. ■

David Schimel is at the National Center for Atmospheric Research, Boulder, Colorado 80305, USA.

The long and the short of it...

Time: how much of the cosmological timescale do we control and use?

Alexander E. Kaplan

A quote often attributed to showman P. T. Barnum is, “a sucker born every minute” — make counting new-born suckers a way of measuring time? A second, rather than a minute, might be closer to the truth — as our heartbeat is nearly a second — but even this scale might still be too short, depending on what there is to measure. Our Universe, according to the ‘big bang’ theory, is about 14 billion years, or 5×10^{17} seconds (s) old. The ultimate timescale (‘Planck time’) of quantum cosmology — approximately 10^{-43} s, the big bang’s birth-flash — is an elementary grain or pixel of time, within which our normal physics of four-dimensional space and time breaks down into much greater number of dimensions as hypothesized by the ‘superstring’ theory.

In logarithmic terms, we, with a lifetime of around 70 years (roughly 2×10^9 s), exist on a scale that has more in common with the age of the Universe than with Planck time. We have learned how to keep track of time — we could even regard ourselves as ‘*Homo temporalis*’ — but how much of it is controlled and used by us? Although the ‘long’ end of this scale is still only of academic interest, the ‘short’ end is becoming a hot and bustling frontier of science and technology. The most familiar examples would be communication and computers. In the quest of higher computer performance, one of the major parameters is the clock frequency or, inversely, the clock cycle. Somewhere in a dark corner of my lab, there are remnants of my old 1989 UNIX computer, which had a clock frequency around 17 MHz. Today’s off-the-shelf computers have a clock cycle of almost 3 GHz, or 0.3 nanosecond (ns, 10^{-9} s).

Lasers have been moving even faster into shorter time domains. Soon after the invention of laser in 1959, the length (duration) of a pulse of light passed the ns and then picosecond (ps, 10^{-12} s) thresholds, and the race was on to get to even shorter pulses. The sub-ps and femtosecond (fs, 10^{-15} s) domain became a field of rich research, with topics such as the registration of super-fast processes, time-resolved spectroscopy, characterization of semiconductors with sub-ps relaxation times, and the control of chemical reactions and fs time-resolution by powerful laser pulses. This domain also hosts the so-called Terahertz technology, which uses these pulses as, for example,

a diagnostic tool to ‘see-through’ opaque materials and structures.

The record for the shortest laser pulses now stands at 4–5 fs, which is close to the length (~ 3 fs) of a cycle of a near-infrared laser. The challenge is to generate controllable pulses even shorter than 1 fs. Why go shorter? The highest spectral frequency of a sub-cycle pulse is inversely proportional to its length, τ . The photon energy is the frequency times the Planck constant, $\hbar$, so we have the pulse’s highest photon energy as $E_{\max} \approx \hbar/\tau$. Although the sub-ps and fs domains correspond to $E \sim 0.1$ – 0.01 eV (which is typical for molecular reactions) the domain below 0.15 fs (150 attoseconds, 150×10^{-18} s) is in atomic physics territory — this is the time it takes an electron at the ground state of a hydrogen atom to revolve around the proton. A few methods to generate such pulses have been proposed. Recently, the sub-fs pulses have been observed experimentally using high-order harmonics. Although these sub-cycle or single-cycle pulses have an extremely broad Fourier spectrum (from radio to extreme ultraviolet domains), they differ dramatically from those generated by regular super-broadband sources (for example, black-body radiation) because ideally all their spectral components have the same phase. Such large-scale trans-spectral coherence has never been encountered in regular optics. Indeed, although super-short pulses are plentiful in the black-body radiation (for example, sunlight), they arrive and behave completely at random. Coherency and controllability make all the difference in the world of pulses.

Beyond the atomic-scale horizon, there are ions of heavy elements. In the ‘ionic extreme’, we can think of the heaviest stable atom, uranium, with all but one electron

away. More than 110 KeV (close to the transition energy of uranium) is needed to eject that last electron, which makes the atomic timescale, $\sim 10^{-20}$ s. Beyond the atomic/ionic physics runs into a quantum desert? Going shorter still, we hit the next domain of fundamental interest: quantum electrodynamics (QED), such as the electron–positron pairs requiring double the rest energy of an electron (1.02 MeV) and strong nuclear reactions, for example, deuterium electro-disintegration producing proton and neutron near 1.2 MeV. These are reminiscent of photoionization in atoms, but on an energy scale up to five orders of magnitude higher. The timescale respectively shrinks to zepto-seconds

(zs, 10^{-21} s). The feasibility of generating and controlling sub-as to zs pulses that may illuminate, time-resolve, and even ultimately control nuclear reactions, have been discussed recently. The idea is to drive free electrons in a tight circle by a laser with the currently available intensity up to 10^{21} W/cm² in the so-called ‘lasertron’ configuration. These electrons, which will be almost instantaneously released and accelerated to the energy $E \sim 50$ MeV in the massive ionization of nanoparticles of matter, should be able to generate photons in QED and nuclear domains.

Past that horizon, we enter the territory of high-energy physics, when charged particles brought to almost the speed of light in huge accelerators collide with target nuclei (or similar counter-propagating particles) to produce a cloud of new elementary particles. If we ever work out how to coherently control the production of the same particles in these collisions, the radiation may be made much faster. A pulse with the highest photon energy of, for example, 1 TeV (millions of MeV) could ideally be $\sim 10^{-27}$ s short. This is still long way from the ultimate time scale, 10^{-43} s...but why bother? Oh, but our curiosity, that ever revving engine of our inquisitive exploration, will most likely make us do it. We are suckers for that... ■

Alexander E. Kaplan is in the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, Maryland 21218, USA.

FURTHER READING

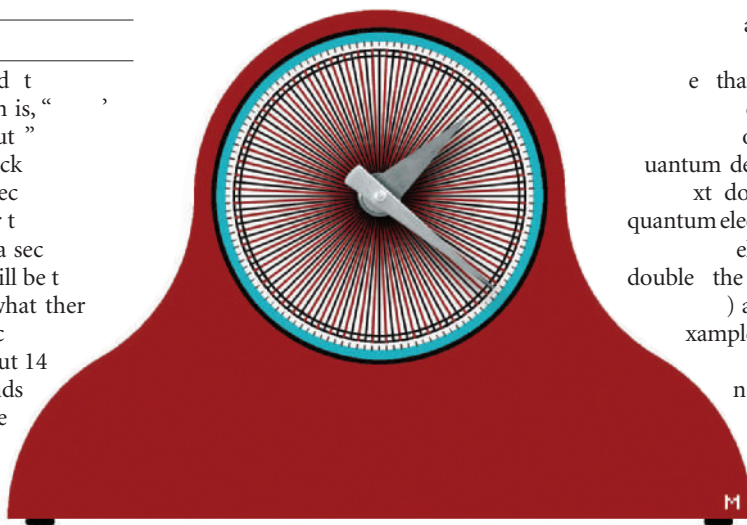
Paul, P. M. *et al. Science* **292**, 1689 (2001).

Hentschel, M. *et al. Nature* **414**, 509 (2001).

Zewail A. *Nature* **412**, 279 (2001).

Kaplan, A. E. & Shkolnikov, P. L. *Phys. Rev. Lett.* **88**, 74801 (2002).

Greene, B. *The Elegant Universe*, (Random House, New York, 2003).



Holding it together in the eye

Paul A. Janmey and Dennis E. Discher

To form tissues, like cells must clump together. The striking resemblance between one cell aggregate in flies and a cluster of soap bubbles points to a crucial role for surface mechanics in biological pattern formation.

How informative are the relatively simple and general rules of physics in explaining the often bewildering complexity and specificity of biology? The large number of proteins and metabolites that work together to produce the specific shapes and functions of different cell types suggests that any apparent similarities between cells and simpler non-biological objects are unlikely to be more than coincidence. However, as Hayashi and Carthew¹ report on page 647 of this issue, there is a remarkable analogy between the structures formed from a type of cell in the retina of fruitfly eyes and clusters of soap bubbles. This analogy illuminates the surface forces that are responsible for producing the specific cellular shapes required for biological function in the eye.

In a developing tissue or organism containing dozens or even thousands of different cell types, the formation of uniquely structured subsets of cells is a complex process. It can involve cell motility, cell adhesion, attractive and repulsive signalling, and other factors, all of which are controlled by their own network of biochemical reactions. However, even before very much was known about the molecular biology underlying any of these events, developmental biologists had noted that simple mixtures of cells become sorted out in a manner that is reminiscent of the behaviour of immiscible liquids². These investigators suggested that similar phenomena, based on minimizing the surface or adhesive energies of cell clusters, might account for cell sorting *in vivo*. Furthermore, *in vitro* studies, which manipulated adhesion strength in systems containing small numbers of different cell types, indicated that concepts of surface energies can predict cell sorting in this context³. Now Hayashi and Carthew¹ show that such predictions also succeed *in vivo*.

The insect eye (Fig. 1) is an attractive model for studying development, because it is a relatively small and simple organ, formed by a limited number of cell types. As structures emerge they can be observed by microscopy, and their formation can be manipulated by precise changes in protein structure, produced by mutations in the genes that code for them. A particularly well-defined and beautiful structure in the insect eye is the ommatidium — the single eye within the compound eye — which contains only around 20 cells. The subject of Hayashi and Carthew's study is a set of cone cells, of

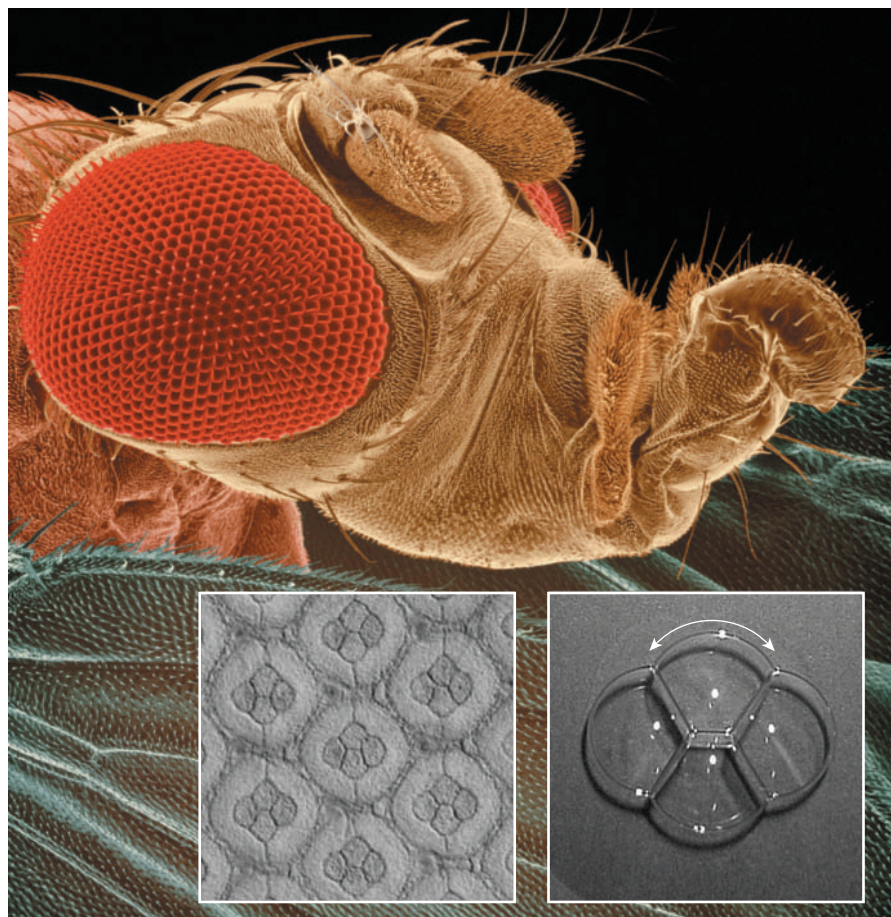


Figure 1 View of a fruitfly's eye (and more). Inset, left, the retina, showing numerous ommatidia and, within each ommatidium, a cluster of four cone cells; right, a cluster of bubbles. The arrow illustrates the presence of tension on the bubble surface. (Inset images taken from ref. 1.)

which there are normally four, but more or less in some mutant flies. These cells lie at the top of the ommatidium, above the light-detecting photoreceptor cells.

Hayashi and Carthew first show that the cone cells configure themselves into a cluster that strikingly resembles a cluster of four soap bubbles (Fig. 1, inset). They then focus on a set of proteins called cadherins, which are responsible for the adherence between cells. They observe differences in the types of cadherins that are expressed by different cells in the ommatidium, noting that one cadherin (N-cadherin) is expressed along the interfaces between individual cone cells. They further find that damaging N-cadherin does not prevent the ommatidium from forming, but does disrupt the typical bubble-like structure of the cluster of cone cells.

Hayashi and Carthew also take advantage of a previously characterized mutant fly in which cell–cell contacts are normal, but the number of cone cells varies from three to six rather than being fixed at the usual four. The kinds of clusters that three to six soap bubbles can make are limited and well characterized, and Hayashi and Carthew find examples of each possible type of cluster in the eyes of these mutant flies. This remarkable correlation suggests that, like clusters of soap bubbles, cone-cell aggregates configure themselves in a way that intrinsically minimizes their overall surface energy.

The idea that aggregates of cells and clusters of soap bubbles take on shapes that minimize surface energy does not, of course, imply that the molecular details of cell adhesion bear much resemblance to those in soap

D. SCHARF/SPL

bubbles. The proteins (such as N-cadherin) and phospholipids that hold one cell membrane to another, and thereby allow the cells to pack in an energy-minimizing way, are very different from the simple surfaces that partition a collection of bubbles.

Something that Hayashi and Carthew do not emphasize, but which can be clearly inferred from their images, is that cell volumes (as well as bubble volumes) remain relatively constant during adhesion — and this might influence the distribution of cell components. Adhesion leads to a surface tension on the non-adherent parts of the membranes⁴ that is evident in the rounded appearance of both bubbles and cells away from the zones of contact. According to Laplace's law, this tension is proportional to a hydrostatic pressure difference across the membrane or interface; but osmotic pressures in cells are comparatively high and resist any large variations in cell volume. So membranes are subject to sustained tensions (something that has only recently been documented), which seem to be linked to various cellular processes⁵. Membrane-channel proteins or cytoskeletal processes that are regulated by tension could therefore be activated locally and feed back to further influence the asymmetric distributions of proteins documented by Hayashi and Carthew.

A surprising aspect of how well the minimization of surface energy accounts for the structures of cone-cell aggregates is that, unlike bubbles, cells are filled not with fluid but with a viscoelastic protein skeleton, both within the cell interior and lining the cell surface. In addition, cells exert forces on each other and on their surrounding extracellular matrix. The internal mechanical properties of retinal cone cells have not been measured directly, but they are unlikely to be those of a simple liquid. And yet these mechanical properties and forces seem — on the long timescales of development — to have little or no influence on the shapes of cell aggregates. This result makes it clear that the laws of physics, which underlie all biological processes, are not always obscured by the dazzling molecular complexity of biology, and that structural similarities between tissues and their simpler inanimate counterparts can be informative about the ways in which biological patterns form. ■

Paul A. Janmey and Dennis E. Discher are at the Institute for Medicine and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.
e-mails: janmey@mail.med.upenn.edu
discher@seas.upenn.edu

- Hayashi, T. & Carthew, R. W. *Nature* **431**, 647–652 (2004).
- Steinberg, M. S. & Poole, T. J. *Phil. Trans. R. Soc. Lond. B* **295**, 451–460 (1981).
- Steinberg, M. S. & Takeichi, M. *Proc. Natl Acad. Sci. USA* **91**, 206–209 (1994).
- Hategan, A., Law, R., Kahn, S. & Discher, D. E. *Biophys. J.* **85**, 2746–2759 (2003).
- Raucher, D. & Sheetz, M. P. *J. Cell Biol.* **148**, 127–136 (2000).

Astromineralogy

Dust in another solar system

Steve Desch

A star surrounded by a disk of dust could be a solar system in the making. Analysis of radiation from the dust suggests that there might be belts of comets or asteroids, and even a planet, orbiting the star.

Astromineralogy is an exciting new field, a combination of astronomy and mineralogy. The goal is to use astronomical observations — usually at mid-infrared (MIR) wavelengths (2–30 μm) — to determine the size, crystal structure and chemical make-up of dust grains in space, often around protostars. Such observations reveal much about the dust distribution in protoplanetary disks, probing their architecture and evolution, and about rocky material in newly forming solar systems — the material that will ultimately form Earth-like planets. Astromineralogical observations are a relatively new achievement, because substantial technological advances in infrared detection were needed. But now the field is coming of age; on page 660 of this issue, Okamoto *et al.*¹ present spatially resolved MIR spectra of the disk around the star β Pictoris (β Pic) that reveal the presence of belts of rocky material, and even the disk's composition. Their observations are a new piece of the puzzle of how solar systems and Earth-like planets form.

The β Pic system is a prime example of a debris disk around a young star in which all of the dust is freshly produced from larger bodies. The dust might have been produced in pulverizing collisions between planetesimals (fledgling planets) or it might have been shed from evaporating comets. The dynamics of the dust grains are determined by the balance between radiation pressure and gravitational forces alone, because there is not an appreciable amount of gas in the β Pic disk². Small dust grains (less than 1.5 μm in radius) are blown out of the disk by β Pic's radiation, on a timescale of tens of years³. Larger dust grains, however, absorb and re-radiate that energy in a way that causes them, over thousands of years, to spiral in towards the star (the Poynting–Robertson effect⁴). As the age of β Pic far exceeds the timescale of both processes — it is 12 million years old⁵ — the dust in its disk must be replenished with freshly delivered debris from comets or planetesimals.

The size, crystal structure and chemical composition of this dust can be determined from observations at MIR wavelengths⁶. Small, warm silicate grains (at temperatures of about 300 K) emit strongly at wavelengths

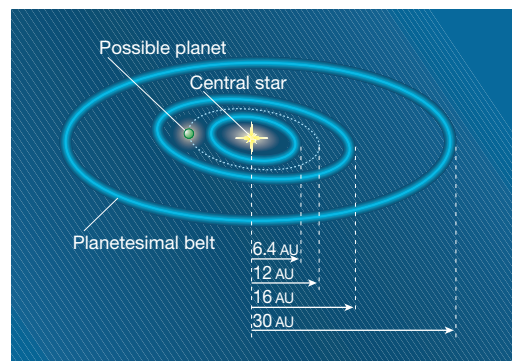


Figure 1 Rings around β Pictoris. Okamoto *et al.*¹ have analysed radiation at mid-infrared wavelengths from the disk of dust that surrounds the star β Pictoris, to decipher the size and mineral composition of the dust grains. The authors suggest that the star is surrounded by three belts that might contain comets or planetesimals (fledgling planets); there might even be a planet in orbit around β Pictoris.

of about 10 μm . Amorphous (non-crystalline) olivines — $(\text{Mg},\text{Fe})_2\text{SiO}_4$ — in particular emit most strongly at 9.7 μm ; the crystalline olivine forsterite, Mg_2SiO_4 , emits strongly at several specific wavelengths, including 11.2 μm . These emission features are most pronounced for small grains, those of radius less than $\lambda/2\pi$, where λ is the wavelength of the radiation. Larger grains tend to emit like black bodies, more continuously across a range of wavelengths. Okamoto *et al.*¹ show that the predicted spectra for dust grains of various sizes and mineralogies — amorphous and crystalline — match their observed dust-emission spectra from the β Pic disk.

Earlier observations have already suggested that the dust in the β Pic disk is concentrated in distinct belts around the star⁷, which implies that the larger bodies that replenish the dust are orbiting at distinct distances. From earlier MIR spectra of the dust observed by Weinberger *et al.*⁸, emission from small silicate grains (less than 10 μm in radius) within a radius of 20 astronomical units around β Pic had been inferred (an astronomical unit, or AU, is the mean Earth–Sun distance). Okamoto *et al.*¹ have improved on the spatial resolution of those observations, from 7.7 to 3.2 AU, and have modelled a variety of specific minerals. Their data show strong emission excesses at a wavelength of 9.7 μm — indicative of small, amorphous silicate grains — at specific radii, of about 6.4, 16 and 30 AU. Large amorphous

silicate grains seem to be concentrated near β Pic (as expected, due to Poynting–Robertson drag).

Crystalline silicate grains are also concentrated near β Pic. This could be attributed to the higher temperatures there, because amorphous silicate grains are annealed (converted to crystalline form) above about 1,100 K (ref. 9). Alternatively, these grains could have been shed by comets entering β Pic's inner solar system¹⁰. Carbon monoxide molecules, which are easily destroyed by ultraviolet radiation from β Pic, have been detected in the β Pic disk³ — strongly implying that they are being replaced through the continuing evaporation of comets. In our own Solar System, a high proportion (about 30%) of the silicate grains shed from comets such as Hale–Bopp are crystalline¹¹.

The presence of small amorphous grains at the particular radii of 16 and 30 AU is intriguing; these radii match the locations of the dust bands previously inferred for the β Pic disk⁷. Okamoto *et al.*¹ have found that the dust grains at these radii are amorphous silicates and are small enough to require constant replenishment. These data strongly suggest the presence of belts of comets or planetesimals at these radii in the β Pic disk (Fig. 1). The improved resolution of Okamoto and colleagues' data, however, has led them further, to the detection of another dust belt at a radius of 6.4 AU. As the authors discuss, this arrangement of a belt of planetesimals or comets at 6.4 AU as well as 16 AU suggests the gravitational influence of a shepherd planet at a radius of 12 AU.

As for the dominance of amorphous silicates in these belts, it could be that there are no crystalline silicates at these radii (unlike the situation inferred from comets from our own Solar System), or that silicates are transformed within the planetesimals from crystalline to amorphous form. Future work can confirm or deny these hypotheses. But it is clear that astromineralogical observations will provide further, exciting insight into the formation of rocky planets in β Pic and other solar systems.

Steve Desch is in the Department of Physics and Astronomy, Arizona State University, Tempe, Arizona 85287-1504, USA.

e-mail: steve.desch@asu.edu

- Okamoto, Y. K. *et al.* *Nature* **431**, 660–663 (2004).
- Lecavelier des Etangs, A. *et al.* *Nature* **412**, 706–708 (2001).
- Artymowicz, P. *Astrophys. J.* **335**, L79–L82 (1988).
- Mukai, T. & Giese, R. H. *Astron. Astrophys.* **131**, 355–363 (1984).
- Zuckerman, B., Song, I., Bessell, M. S. & Webb, R. A. *Astrophys. J.* **562**, L87–L90 (2001).
- Molster, F. J. & Waters, L. B. F. M. in *Astromineralogy (Lecture Notes in Physics no. 609)* (ed. Henning, T. K.) 121–170 (Springer, Berlin, 2003).
- Wahhaj, Z. *et al.* *Astrophys. J.* **584**, L27–L31 (2003).
- Weinberger, A. J., Becklin, E. E. & Zuckerman, B. *Astrophys. J.* **584**, L33–L37 (2003).
- Hallenbeck, S. L., Nuth, J. A. III & Nelson, R. N. *Astrophys. J.* **535**, 247–255 (2000).
- Li, A. & Greenberg, M. *Astron. Astrophys.* **331**, 291–313 (1998).
- Wooden, D. *et al.* *Astrophys. J.* **517**, 1034–1058 (1999).

Molecular biology

No exception to reversibility

Yi Zhang

Histone proteins, which serve as scaffolds for packaging DNA, can be modified in numerous ways. It's been thought that one modification, methylation, is irreversible — but that view must now change.

Each of our cells contains about two metres or so of DNA, which must be packed down very tightly to fit into the cell nucleus. The compact form of DNA is known as chromatin, the basic unit of which consists of DNA wrapped around an octamer of histone proteins¹. The histones are not merely architectural proteins, however: they also influence chromatin dynamics. One way in which they do so is through their covalent modification with certain chemical groups or small proteins — acetyl groups, phosphate groups, ubiquitin proteins or methyl groups². Enzymes that catalyse the addition or removal of the first three modifications have been identified.

But enzymes that remove methyl groups have been more elusive, raising the question of whether methylation is an exception to the rule: the only histone modification that is irreversible³. Two new papers, however — published in *Science* by Wang *et al.*⁴ and in *Cell* by Cuthbert *et al.*⁵ — reverse that view.

The histone octamer consists of two copies each of four 'core' histones: H2A, H2B, H3 and H4. Methylation occurs on certain lysine and arginine amino acids within H3 and H4, and is catalysed by distinct families of enzymes⁶; PRMT1 and CARM1, for instance, are enzymes involved in arginine methylation. Lysine residues can be mono-, di- or tri-methylated, whereas arginine

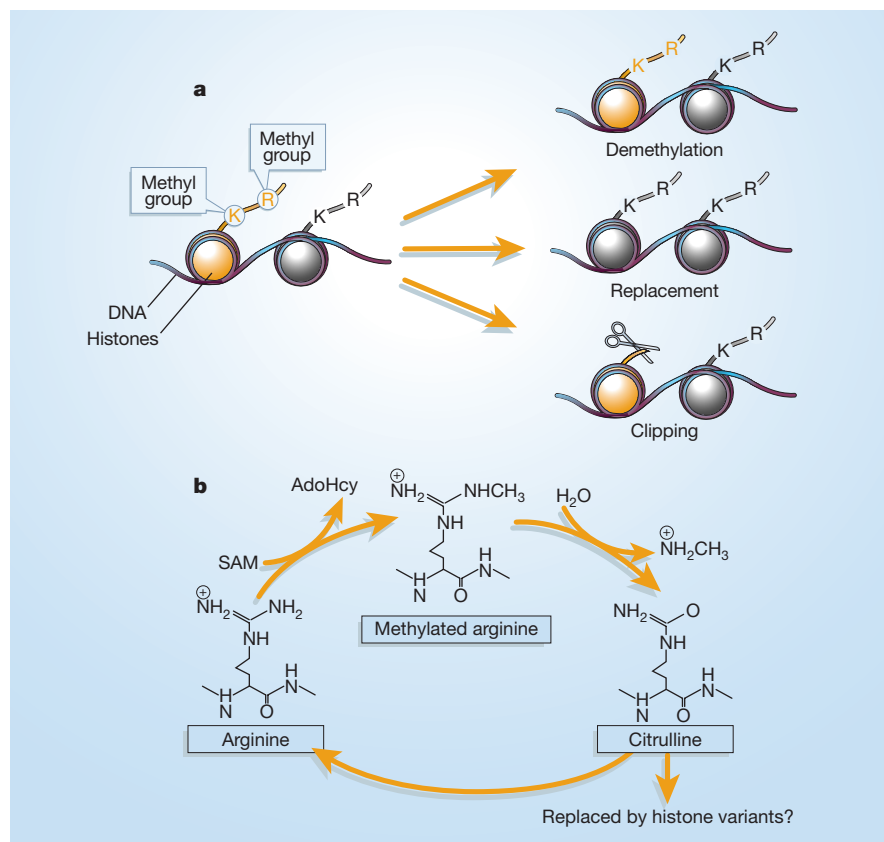


Figure 1 Reversing methylation in histone proteins. a, Three theoretical mechanisms: top, enzymatic demethylation, perhaps regulated by the HDM enzyme; centre, replacement of methylated histones by unmethylated variants; bottom, clipping of histone 'tails', by enzymes unknown. K, lysine; R, arginine. b, Wang *et al.*⁴ and Cuthbert *et al.*⁵ have discovered another mechanism: they show that methylated and unmethylated arginine amino acids in two histone proteins can be converted to citrulline. A possible cycle of events is shown. Methylation of arginines is accomplished by the PRMT1 or CARM1 enzymes, with SAM (S-adenosyl-L-methionine) as cofactor; AdoHcy (S-adenosyl-L-homocysteine) is released. Conversion to citrulline is achieved by PAD4/PAD14. It remains unknown what then happens to citrulline — whether it is converted back to arginine (by an aminotransferase enzyme, perhaps), or whether histones containing citrulline are replaced by unmodified versions.



100 YEARS AGO

The recent tube operations in London have brought to the surface specimens of the London Clay from different districts. Samples of this clay taken from such different points as Hyde Park Corner, Brompton Road, and Haverstock Hill have been tested in the physical laboratory of the South-western Polytechnic for the presence of a radio-active gas by Mr. H. Cottam, and he has been unable to detect with his apparatus any marked quantity of active gas from the clays. With the same apparatus he has detected quite easily the radio-active gas from the water of a deep well... which goes below the clay to the greensand. We have come to the conclusion that the London Clay forms a floor through which the radio-active gas does not penetrate; or it may be said that the radio-active substance only travels when the water with which it is associated can travel. This is an argument in support of Prof. J. J. Thomson's view, that the radio-active gas, which he found in deep well waters, arises from the splitting up of a trace of soluble radium salt which comes up with the water.

From *Nature* 6 October 1904.

50 YEARS AGO

Little more than a hundred years have passed since hunters such as Gordon Cumming were at their work among the vast herds of big game in the southern quarter of Africa, and little more than fifty since Selous was making a living by ivory hunting somewhat farther to the north. All those countless thousands of animals have now disappeared for ever... The opening-up and white settlement of East Africa did not come until later, but although the same process of decimation began, and many parts of the country have been cleared of game, it is still possible to see great herds of wild animals that recall the accounts of conditions in South Africa given by the early travellers. Nevertheless, the presence of enormous herds of game animals is quite incompatible with the economic exploitation of the country and the rapid expansion of the native population; the game must go, and there can be no hope of its survival outside the National Parks and game reserves... As yet, however, there is practically no information available on the biology of these mammals, information that is essential for successfully managing such parks and reserves.

From *Nature* 9 October 1954.

residues can only be mono- or di-methylated.

Histone methylation has been linked to biological processes ranging from the regulation of gene transcription, to the inactivation of one copy of the X chromosome in females, to RNA-mediated gene silencing^{2,6,7}. One way in which it works is by serving as a docking site for other proteins². The nature of a specific methylation determines the protein that it recruits, which in turn dictates the biological outcome (see, for example, ref. 6).

Unlike other histone modifications, methylation has generally been regarded as stable³ — a notion that comes from early studies showing that histone proteins and methylated lysine or arginine residues within them have similar turnover rates⁸. But although a non-reversible methyl mark would fit with a role for histone methylation in long-term gene silencing, it is not compatible with situations in which rapid reversal of gene expression takes place. To solve this paradox, mechanisms including enzyme-catalysed demethylation, replacement of methylated histones by unmodified histones, and clipping of methylated histone 'tails' have been proposed^{3,9} (Fig. 1a) — but none has yet been demonstrated experimentally. Now, however, Wang *et al.*⁴ and Cuthbert *et al.*⁵ have found that the human enzyme peptidylarginine deiminase 4 (PAD4/PADI4) can catalyse the conversion of methylated arginines to citrulline, providing yet another mechanism by which histone methylation levels could be controlled (Fig. 1b).

How is it that two groups independently identified the same enzyme and came to similar conclusions? Previous studies¹⁰ established that arginine residues in other proteins can be converted to citrulline by enzymes of the peptidylarginine deiminase family. Of this family, only PAD4/PADI4 is found in the nucleus; moreover, its expression correlates with the appearance of citrulline in histones¹¹. These facts made it a good candidate for a histone arginine demethylase.

So Wang *et al.* and Cuthbert *et al.* carried out direct tests of PAD4/PADI4 *in vitro*. They found that it 'deiminated' numerous unmethylated arginines — converted them to citrulline — in histones H3 and H4. It also decreased the amount of arginine methylation that is catalysed by PRMT1 and CARM1 in both histones. Notably, at the same time that the amount of this methylation decreased, the quantities of citrulline in both histones increased. The detection of methylamine as a released product⁴ supports the idea that methylated arginine is a genuine substrate of this enzyme. So, PAD4/PADI4 can catalyse both the deimination of unmethylated arginine and the 'demethylation' of methylated arginine *in vitro*. As Wang *et al.* find, it can also do so in granulocyte cells.

What are the consequences of these

reactions? First, as Cuthbert *et al.* show, the conversion of histone arginines to citrullines actually prevents histone methylation by CARM1. Second, Wang *et al.* find that demethylation of methylated histone arginines reverses the effects associated with methylation. So, PAD4/PADI4 presumably antagonizes the functions of the methylating enzymes CARM1 and PRMT1. For instance, previous studies have linked histone arginine methylation by these enzymes to transcriptional activation by nuclear hormone receptors^{12–15} — so PAD4/PADI4 is likely to repress such transcription. Indeed, both groups link the recruitment of PAD4/PADI4 to an oestrogen-responsive gene, pS2, with the appearance of citrullinated histones and the downregulation of this gene.

These papers^{4,5} provide convincing evidence that the methylation of arginine amino acids in histone proteins can be reversed enzymatically. But, as ever, the findings raise questions. For example, it seems that PAD4/PADI4 has a very loose substrate specificity *in vitro*. It works on both methylated and unmethylated arginines. And the arginine residues it deiminates are not limited to the sites that are targeted by CARM1 and PRMT1; indeed, arginine 8 in histone H3, a site not known to be methylated by either enzyme, is the preferred target⁴. There might be an interplay between the citrullination of H3 arginine 8 and the methylation of H3 lysine 9, but for now the significance of this citrullination remains unknown.

The second issue worth noting is that, *in vitro*, PAD4/PADI4 cannot demethylate H4 or H3 peptides containing di-methylated arginines^{4,5}. This presents a puzzle. Although mass-spectrometry analysis^{14,15} identified mono-methyl groups as the major methyl form on arginine 3 in histone H4, most arginines 17 and 26 in histone H3 become di-methylated after incubation with CARM1 *in vitro*¹⁶. How, then, can this methylation be removed? Although an unidentified enzyme might be required, the available evidence suggests that PAD4/PADI4 is responsible: for instance, in granulocytes, activation of this enzyme led to less methylation on H4 arginine 3 and H3 arginine 17, as analysed using site-specific antibodies that recognize di-methylated arginine⁴. The most likely solution is that PAD4/PADI4 can work only on intact chromatin substrates. Alternatively, it might need to work with a partner.

As well as revealing that arginine methylation can be reversed, the new papers^{4,5} show that a new form of histone proteins — containing citrulline residues — exists in cells. This finding, too, raises questions. How stable are citrullinated histones? Do they have any effects on chromatin structure? Can other proteins recognize and bind to them? And what is the fate of these histones? With regard to the last question, the transient presence of citrullines in the pS2 gene⁵ suggests

that their rapid removal must be possible. The availability of site-specific antibodies against citrullinated histones will allow at least some of these issues to be addressed. And one final question: is the methylation of lysine residues in histone proteins also reversible? ■

Yi Zhang is in the Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7295, USA.
e-mail: yi_zhang@med.unc.edu

1. Kornberg, R. D. & Lorch, Y. *Cell* **98**, 285–294 (1999).
2. Jenuwein, T. & Allis, C. D. *Science* **293**, 1074–1080 (2001).
3. Bannister, A. J., Schneider, R. & Kouzarides, T. *Cell* **109**, 801–806 (2002).

4. Wang, Y. *et al. Science* doi:10.1126/science.1101400 (2004).
5. Cuthbert, G. L. *et al. Cell* **118**, 545–553 (2004).
6. Zhang, Y. & Reinberg, D. *Genes Dev.* **15**, 2343–2360 (2001).
7. Grewal, S. I. & Moazed, D. *Science* **301**, 798–802 (2003).
8. Byvoet, P. *Arch. Biochem. Biophys.* **152**, 887–888 (1972).
9. Henikoff, S., Furuyama, T. & Ahmad, K. *Trends Genet.* **20**, 320–326 (2004).
10. Vossenaar, E. R., Zendman, A. J., van Venrooij, W. J. & Pruijn, G. J. *BioEssays* **25**, 1106–1118 (2003).
11. Nakashima, K., Hagiwara, T. & Yamada, M. *J. Biol. Chem.* **277**, 49562–49568 (2002).
12. Bauer, U. M., Daujat, S., Nielsen, S. J., Nightingale, K. & Kouzarides, T. *EMBO Rep.* **3**, 39–44 (2002).
13. Chen, D. *et al. Science* **284**, 2174–2177 (1999).
14. Strahl, B. D. *et al. Curr. Biol.* **11**, 996–1000 (2001).
15. Wang, H. *et al. Science* **293**, 853–857 (2001).
16. Schurter, B. T. *et al. Biochemistry* **40**, 5747–5756 (2001).

Palaeontology

Ecology of ice-age extinctions

John Pastor and Ron A. Moen

The last ice age saw the extinction of numerous large mammals — but perhaps not as many as was thought. The woolly mammoth survived to much more recent times, and so, it now seems, did the Irish elk.

Sabre-toothed tigers, mastodons, woolly mammoths — these and many other spectacular large mammals are generally thought to have become extinct about 10,000 years ago, at the end of the Pleistocene epoch, otherwise known as the last ice age. But it's becoming clear that some of these species clung on tantalizingly close to the present day. Thomas Jefferson's instruction to Meriwether Lewis and William Clark to search for live woolly mammoths in the American West in 1804 was perhaps a little optimistic. But the species survived on Wrangel Island in the northeastern Siberian Arctic until some 4,000 years ago¹, making it contemporaneous with the Bronze Age Xia Dynasty in China. On page 684 of this issue, Stuart *et al.*² report that another charismatic ice-age mammal that was thought to have become extinct 10,000 years ago — the giant deer or Irish elk (*Megaloceros giganteus*) — survived in western Siberia to the dawn of historic times. The finding lends weight to the idea that there is no one explanation for the so-called Pleistocene extinctions.

The Irish elk (Fig. 1) must have cut an impressive figure, standing more than two metres high at the shoulder — about the same as a bull moose, the largest living member of the deer family. But when and why did it become extinct? In their investigation, Stuart *et al.*² began by carrying out radiocarbon dating of five skeletal specimens, including a complete skeleton of an antler-bearing male. By combining this information with maps of the specimens' locations, they show that Irish elk were widespread in Europe — from Ireland to Russia, and from Scandinavia to the Mediterranean — before 20,000 years ago.

But by the last glacial maximum 15,000 years ago, they may have been restricted to refuges in the shrub steppes of central Asia. From there, Irish elk apparently recolonized north-western Europe following the retreat of the Alpine and Scandinavian ice sheets during a period of climatic warming. The European

population made a last stand in the British Isles before dying out 10,500 years ago, but the Siberian population persisted for another 3,000 years.

What caused the extinction of so many large mammals 10,000 or so years ago? Human hunting³, changes in climate or vegetation, or both⁴, are often proposed to be causal factors. But the “ragged” nature (to use Stuart and colleagues' phrase) of these Late Pleistocene extinctions, with isolated pockets of populations surviving for longer, suggests that the extinctions have a complex ecology, with no single mechanism responsible for the demise of every species in every location.

Theories for both the expansion and the extinction of Irish elk populations, for instance, often focus on the animals' huge antlers, which weighed 40 kilograms and spanned 3.5 metres, making them 30% larger than those of modern moose. It has been suggested⁵ that female Irish elk selected males with large antlers, as this might have signified an ability to find sufficient food to support building and shedding a rack each year. This ability would then be passed on to their male progeny.

But the large antlers, which contained as much as 8 kilograms of calcium and 4 kilograms of phosphate, would have posed a large annual nutritional burden on bulls⁶. The antlers would also be physically unwieldy in dense forests⁷. So both physical and nutritional constraints probably restricted the



Figure 1 Prehistoric giant — artist's impression of the Irish elk.

Irish elk to productive open environments, with relatively tall willow and birch shrubs that could be navigated but still supply sufficient calcium and phosphate for antler growth^{6,7}. An inability to balance sexual selection for large antlers with nutritional selection pressures for smaller antlers may have led to the Irish elk's demise in the British Isles, particularly as the climate cooled rapidly and caused the vegetation to change to short-statured and unproductive tundra⁶.

And what about conditions in Siberia, where the Irish elk evidently survived for some while longer before its eventual demise? Stuart *et al.* suggest that the necessary productive and open grass-shrub plant community persisted in the eastern foothills of the Urals, sandwiched between dense forests on the upper mountain slopes and dry grass steppes in the plains to the east (neither of which could have supported the nutritional demands of the large antlers). The authors also propose that a general spread of the dry grass steppes over western Siberia, and the persistence of closed forests in the mountains, squeezed the grass-shrub habitat to a sliver and contributed to the demise of perhaps the last population of Irish elk about 7,000 years ago. In contrast, woolly mammoths survived on the steppe-tundra of Wrangel Island well beyond the extinction of the Irish elk. Stuart *et al.* effectively link the appearance and disappearance of Irish elk and woolly mammoths in the fossil records to palaeobotanical data showing climate-driven changes in the vegetation communities that are thought to have supported these large herbivores.

In recent years, ecologists have shown that large mammalian herbivores not only respond to changes in the composition of plant communities, but also themselves control this composition and plant productivity, through nutrient redistribution and selective foraging on preferred plant species^{8,9}. Intriguingly, by trampling tundra mosses and stimulating grass growth through grazing, the woolly mammoth might have helped to maintain and expand its own preferred dry steppe habitat, even when the climate was cool enough to support a mossy tundra¹⁰. The shift to mossy tundra in Siberia might have happened only after human hunters extirpated woolly mammoths, thus relieving both trampling pressure on mosses and competition with grasses that were previously stimulated by grazing¹⁰. The Irish elk might also have affected its own habitat through browsing and grazing, but this remains an open question.

So the causes of the Late Pleistocene extinctions of large mammals are complex. Climate warming or cooling may have set the stage through broad-scale vegetation changes, and human hunting may have delivered the *coup de grâce*. But the animals themselves contended with pressures from

nutritional and sexual selection, exacerbated perhaps by how they and their competitors affected their own environments. Stuart and colleagues' work exemplifies the wealth of detail that it is possible to obtain on the spatial and temporal patterns of fossil distribution. This, together with the finely resolved records of vegetation change generated by fossil pollen studies, and the development of realistic simulations of the physiology and ecology of extant and extinct large mammalian herbivores^{6,10}, suggests that the time is ripe for a more comprehensive analysis of the ecology of the Late Pleistocene extinctions. ■

John Pastor and Ron A. Moen are at the University

of Minnesota, Duluth, Minnesota 55811, USA.

e-mail: jpastor@nrri.umn.edu

1. Vartanyan, S. L., Garrut, V. E. & Sher, A. V. *Nature* **362**, 337–340 (1993).
2. Stuart, A. J., Kosintsev, P. A., Higham, T. F. G. & Lister, A. M. *Nature* **431**, 684–689 (2004).
3. Martin, P. S. in *Quaternary Extinctions: A Prehistoric Revolution* (eds Martin, P. S. & Klein, R. G.) 364–403 (Univ. Arizona Press, Tucson, 1984).
4. Stuart, A. J. *Biol. Rev.* **66**, 453–562 (1991).
5. Geist, V. *Nat. Hist.* **95**, 54–65 (1986).
6. Moen, R. A., Pastor, J. & Cohen, Y. *Evol. Ecol. Res.* **1**, 235–249 (1999).
7. Gould, S. J. *Evolution* **28**, 191–220 (1974).
8. Hobbs, N. T. *J. Wildl. Mgmt* **60**, 695–713 (1996).
9. Jefferies, R. L., Klein, D. R. & Shaver, G. R. *Oikos* **71**, 193–206 (1994).
10. Zimov, S. A. *et al. Am. Nat.* **146**, 765–794 (1995).

Materials physics

Doping control for nanotubes

Reshef Tenne

Adding guest atoms to inorganic nanotubes, known as 'doping', influences their room-temperature magnetic properties — properties that could be exploited in 'spintronic' devices and computer memory.

Much of the technology underlying our computers and electronic devices is based on the transport of electronic charge across sub-micrometre structures made from perfectly crystalline silicon. The dimensions of silicon-based devices, transistors in particular, are constantly shrinking (90-nm transistor gates are now in production) and their integration densities are increasing — in line with Moore's law, which predicts at least a doubling of the transistor density on a chip every two years. Although this trend has been followed for almost 40 years, the rate of increase is expected to level off in the near future,

when transistor size reaches a few tens of nanometres.

This intrinsic limitation of silicon devices has provoked much speculation about the 'next big thing' in information processing — might it be 'spintronics', which uses an electron's direction of magnetization, or spin, instead of its charge as the information carrier? Each spin could carry a bit of information — in effect, a single-spin transistor — which could lead to faster computers that consume less electricity, and alleviating the problem of heat dissipation. In designing spintronic devices, we can explore the exciting opportunities offered

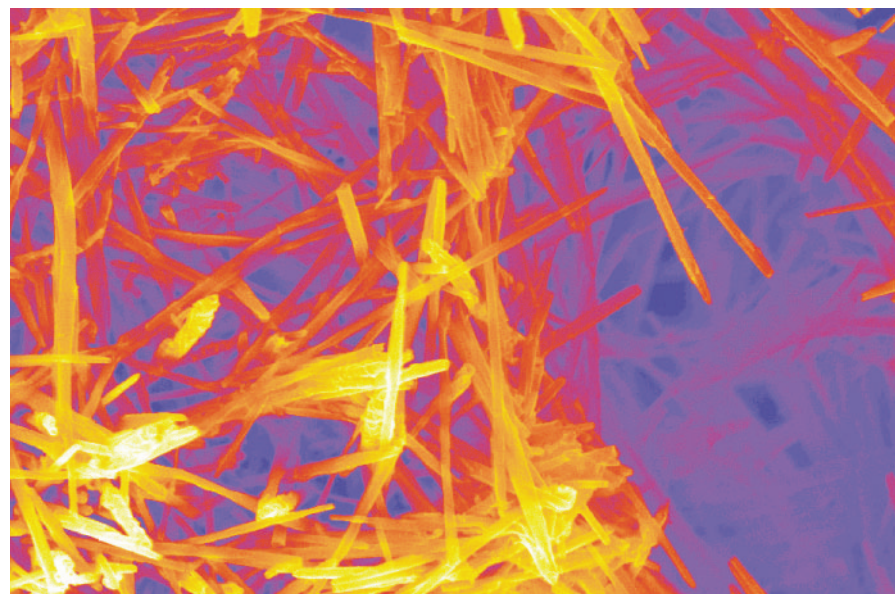


Figure 1 Vanadium oxide (VO_x) nanotubes, seen in false-colour by scanning electron microscope. Krusin-Elbaum *et al.*¹ show how the magnetic properties of such nanotubes can be engineered through doping with guest atoms.

by the chemistry of nanotubes: Krusin-Elbaum *et al.*¹ lead the way, on page 672 of this issue.

Spintronics requires the fabrication of ferromagnetic nanostructures that, at room temperature, can transport spin-polarized carriers, and which can be assembled into addressable hierarchies on a macroscopic chip. Most efforts have been directed towards the mixing of transition-metal atoms (such as Ni, Fe and Mn, which have permanent magnetic moments) into semiconductor devices based on compounds from groups II–VI (such as CdS) or III–V (GaAs) of the periodic table. Superstructures consisting of alternating ferromagnetic/diamagnetic, metallic/oxide thin films have also received attention; like spin valves, spin-polarized currents can be injected into them and transported.

Shortly after the discovery of carbon nanotubes, it was suggested that the layered structures of some inorganic compounds (such as WS₂, MoS₂ and BN; refs 2–4, respectively) should also fold in on themselves and close up into fullerene-like nanostructures and nanotubes. Various strategies have been developed to synthesize such nanotubes. These can be grossly divided into high-temperature processes and ‘chemie douce’ (soft chemistry), a generic name used for wet synthetic processes that are limited to temperatures of 300 °C and below. Among the early examples of chemie-douce synthesis were titanate^{5,6}, niobate⁷ and vanadium-oxide (VO_x) nanotubes⁸ — the last of these being the type of nanotube used by Krusin-Elbaum *et al.*¹ in their study (Figs 1, 2).

To create nanostructures exhibiting room-temperature ferromagnetism, the authors¹ synthesized VO_x-alkylamine nanotubes through a well-established procedure (a combination of sol–gel and hydrothermal synthesis⁸). They then ‘doped’ the nanotubes, introducing foreign atoms into the gaps between layers: lithium atoms, which act as electron donors to the vanadium lattice; and iodine atoms, which act as electron acceptors, creating positive charge carriers in the lattice called ‘holes’. In contrast to the transition-metal devices mentioned above (in which a 3d transition-metal atom is added into a non-magnetic host), here both guest atoms are non-magnetic but the host is magnetized by the elimination or addition of one extra charge (positive or negative).

Careful analysis of the data has led Krusin-Elbaum *et al.*¹ to conclude that the undoped nanotube is a frustrated spin liquid, meaning that its electron spins have random orientations. Although they have not clarified the exact location of the dopant atoms in the nanotube lattice, the authors found that when one charge is added to or eliminated from the highest-energy vanadium atom, the frustration is removed. There is then antiferromagnetic alignment

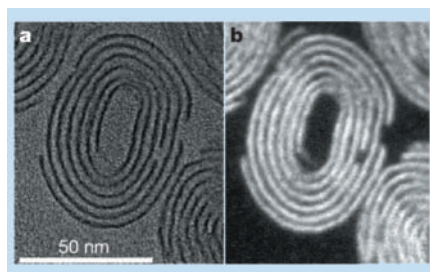


Figure 2 Nano-scrolls. These images from a transmission electron microscope show: a, a VO_x nanotube in cross-section; b, with energy filtering.

of the vanadium spins in the lattice of the nanotube — that is, neighbouring spins have opposite orientations. If a small magnetic field is applied, the spins flip such that they all have the same orientation; this is ferromagnetic ordering. Furthermore, the undoped nanotube is an insulator (a Mott insulator), but the doping transforms it into a good conductor, by shifting the Fermi level (the surface defined by the highest occupied electron energy states) into the conduction or valence bands (for lithium or iodine atoms, respectively). So the doping process turns the VO_x nanotubes into room-temperature magnets that are good charge carriers, accomplishing two of the main requirements for spintronics technology.

The room-temperature magnetization is, however, only modest. But this novel strategy¹ paves the way for creating nanostructures with larger magnetic effects. Furthermore, in analogy to the ‘ballistic’ transport of electrons in single-wall carbon nanotubes,

higher carrier mobilities can be expected if the crystalline order of the nanotubes is improved and if the doping atoms are more carefully controlled. The bipolar nature of the nanotubes — the possibility of making them conductors of either electrons or holes through appropriate doping — is important for the fabrication of transistor structures and junctions (p–n junctions).

Driven by scientific curiosity as well as by the possibilities of application, the burgeoning field of nanotube research benefits from the diversity of inorganic compounds from which the tubes can be made and the fact that their properties can be modified and controlled through judicious chemical transformation, as Krusin-Elbaum *et al.*¹ have demonstrated. Potential applications are numerous — nanotubes as solid lubricants (based on fullerene-like WS₂ and MoS₂ nanoparticles⁹) are already heading for large-scale commercialization. Spintronics may well be the ‘next big thing’.

Reshef Tenne is in the Department of Materials and Interfaces, Weizmann Institute of Science, Rehovot 76100, Israel.

e-mail: reshef.tenne@weizmann.ac.il

1. Krusin-Elbaum, L. *et al.* *Nature* **431**, 672–676 (2004).
2. Tenne, R., Margulis, L., Genut, M. & Hodes, G. *Nature* **360**, 444–446 (1992).
3. Margulis, L., Salitra, G., Tenne, R. & Talianker, M. *Nature* **365**, 113–114 (1993).
4. Chopra, N. G. *et al.* *Science* **269**, 966–967 (1995).
5. Kasuga, T., Hiramatsu, M., Hason, A., Sekino, T. & Niihara, K. *Langmuir* **14**, 3160–3163 (1998).
6. Chen, Q., Zhou, W. & Peng, L.-M. *Adv. Mater.* **14**, 1208–1211 (2002).
7. Saupe, G. B. *et al.* *Chem. Mater.* **12**, 1556–1562 (2000).
8. Spahr, M. E. *et al.* *Angew. Chem. Int. Edn* **37**, 1263–1265 (1998).
9. Rapoport, L. *et al.* *Nature* **387**, 791–793 (1997).

Cell biology

Light on pits

Elizabeth Smythe

The imaging of events in living cells offers a way to test models of cell behaviour and develop new hypotheses. The invaginating ‘pits’ on the surface of cells are the latest subject of this approach.

Writing last month in *Cell*, Ehrlich and colleagues¹ reported progress in a 40-year line of investigation. They have studied the ‘coated pits’ that form on the surface of cells and which take up material from the surface and the extracellular environment.

Seminal work by Roth and Porter² first showed that there are regions on the surface of mosquito eggs that are specialized to allow these cells to take up yolk proteins from their environment. Roth and Porter called these areas coated pits because the surface facing the inside of the cell was coated with an electron-dense material. That material was later shown to be made up of a protein named clathrin³. Over the past 20 years, many other components of coated pits have been

identified. These include the AP2 adaptor protein, which links transmembrane cargo molecules to the clathrin lattice, and dynamin, an enzyme required to transform coated pits into small, spherical vesicles inside the cell⁴ (although its precise role remains controversial).

Clathrin-coated pits have become firmly established as the major pathway for the uptake of nutrients and signalling molecules in almost all non-bacterial cells. This pathway is also essential in tissue-specific processes such as immune surveillance, and it is frequently hijacked by pathogens seeking entry into the cell⁴. A combination of biochemistry, electron microscopy and cell biology has provided a series of snapshots of the coated-vesicle cycle. These snapshots define

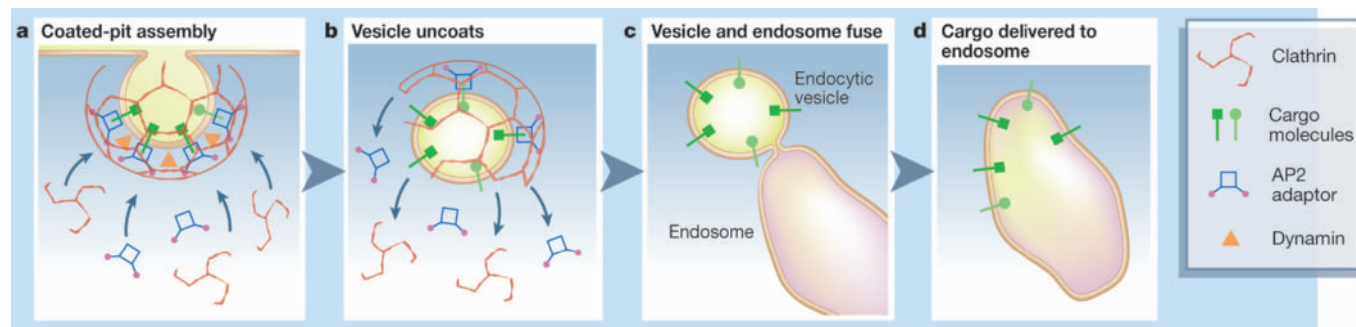


Figure 1 Coated pits and cargo molecules. **a**, Coated pits form following the recruitment of clathrin, AP2 and dynamin (and also other components that are not shown) from the cytoplasm to the inner surface of the plasma membrane. Cargo molecules, which include transmembrane receptors, are

brought into the pits, which become increasingly invaginated and eventually pinch off, forming coated vesicles. **b**, Clathrin and AP2 are rapidly removed. **c, d**, The resulting 'endocytic' vesicle can fuse with — thereby delivering its cargo to — the early endosome.

a temporal order of events (Fig. 1) that includes the assembly of clathrin and other coat proteins just under the cell surface, the selection of cargo molecules, an increase in membrane curvature, and scission of the invaginating pit. The vesicle thereby produced then rapidly loses its coat, allowing it to fuse with another intracellular compartment, the early endosome.

By labelling clathrin, AP2 and cargo molecules with fluorescent 'tags', Ehrlich *et al.*¹ have now gone a step further: rather than taking static snapshots, they have produced movies of the vesicle cycle in living cells. This is not the first example of live-cell imaging of coated-pit formation⁵. But it is the first such study to probe the connection between coat proteins, such as clathrin and AP2, and cargo molecules (such as low-density lipoprotein and its receptor) that are known to be internalized by this route. They thus provide compelling evidence that the fluorescent particles observed indeed represent functional coated pits. The authors introduced the fluorescently tagged molecules into cells cultured under conditions in which the molecules were stably expressed, and subsequently identified growing coated pits through an increase in the size of fluorescent spots on the plasma membrane with time.

It has been suggested⁶ that there are 'hotspots' on the inside of the cell surface at which coated pits assemble, implying that initiator components are specifically localized to such hotspots. But Ehrlich *et al.* observe that the initiation of new pits occurs randomly, suggesting that the proteins or lipids responsible must also be distributed randomly on the plasma membrane.

The authors also find that the fluorescently labelled cargo is recruited to newly forming pits. Surprisingly, though, a significant proportion of the developing pits (around 25%) rapidly disassemble without incorporating cargo. So Ehrlich *et al.* suggest that the capture of cargo might stabilize nascent pits and commit them to forming vesicles. In practice, this would prevent cells from assembling vesicles in the absence of cargo, thus efficiently coupling function to

assembly. The idea that cargo recruitment is linked to successful vesicle formation is consistent with the observation that cargoes such as growth-factor receptors can modulate the vesicle machinery to facilitate their own internalization⁷. Recruitment of such cargo early in pit formation would provide a window of opportunity where the rate at which pits commit to forming vesicles might be enhanced.

Ongoing debate surrounds the role of dynamin, which has been implicated in the final pinching-off of coated pits. Ehrlich and colleagues' observation that dynamin is recruited to newly forming coated pits at early time points is consistent with studies that show a role for this protein in pit invagination as well as scission⁸. However, as dynamin was also detected on the pits that disassemble prematurely, it is unlikely to be a major player in the stabilization of nascent pits. If there is instead a link between cargo recruitment and commitment to vesicle formation, coat components such as epsin, which has been implicated both in the generation of membrane curvature⁹ and as a possible cargo-binding protein¹⁰, might mediate this coupling.

Although this is a pleasing interpretation, other possibilities exist, not least because the coated pits that abort might be expected to be able to capture, and be stabilized by, other, unlabelled cargo molecules, which are relatively abundant on the cell surface. One alternative possibility is that a 'commitment' step directly precedes, and is necessary for, subsequent cargo recruitment. In this model, cargo recruitment and a commitment to vesicle formation would still be linked, but the commitment step would come first.

The authors also find that relatively large cargo molecules, such as reovirus, take longer than smaller cargoes to become included in a fully formed coated pit, hinting that the type of cargo might help to determine pit size. Indeed, hen egg cells have been found to contain large coated pits that are specialized for the uptake of yolk proteins. But generally, coated pits contain more than one type of cargo molecule and, when

visualized by electron microscopy, have a fairly uniform diameter (100–200 nm), which should be large enough to accommodate a reovirus particle. To determine whether clathrin assembles at the same rate around cargoes of different sizes (thus leading to coated vesicles of different sizes), or whether different cargoes affect the rate of coat assembly, pit formation rates in the presence of large and small cargoes will need to be compared under the same conditions.

Ehrlich *et al.* also examined the dynamics of other fluorescently tagged components of coated pits. They find that AP2 and clathrin assemble into coated pits and disassemble from coated vesicles at apparently the same rates as each other. *In vitro* data have shown that assembly and disassembly of the two molecules have very different biochemical requirements; in fact, they have even been shown to be uncoupled in cells in which clathrin is knocked out¹¹. Ehrlich and colleagues' findings show that these processes are nonetheless coordinated with precision *in vivo*.

The exciting aspect of this study¹ is that it allows cell biologists to re-evaluate their models of the coated-vesicle cycle, which are based on a huge body of biochemical and cell-biological data. Furthermore, Ehrlich and colleagues' approach should make it possible to examine the behaviour of other coated-pit components — and the technique could be adapted to reveal what happens when key proteins are mutated or ablated in live cells.

Elizabeth Smythe is in the Department of Biomedical Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.
e-mail: e.smythe@sheffield.ac.uk

1. Ehrlich, M. *et al.* *Cell* **118**, 591–605 (2004).
2. Roth, T. F. & Porter, K. R. *J. Cell Biol.* **20**, 313–323 (1964).
3. Pearse, B. M. *Proc. Natl Acad. Sci. USA* **73**, 1255–1259 (1976).
4. Conner, S. D. & Schmid, S. L. *Nature* **422**, 37–42 (2003).
5. Rappoport, J., Simon, S. & Benmerah, A. *Traffic* **5**, 327–337 (2004).
6. Gaidarov, I., Santini, F., Warren, R. A. & Keen, J. H. *Nature Cell Biol.* **1**, 1–7 (1999).
7. Clague, M. J. & Urbe, S. *J. Cell Sci.* **114**, 3075–3081 (2001).
8. Hill, E. *et al.* *J. Cell Biol.* **152**, 309–323 (2001).
9. Ford, M. G. *et al.* *Nature* **419**, 361–366 (2002).
10. Traub, L. M. *J. Cell Biol.* **163**, 203–208 (2003).
11. Wettestad, E. R. *et al.* *Science* **297**, 1521–1525 (2002).

Cell biology

Nuclear squeeze and stretch

J. Cell Sci. **117**, 4779–4786 (2004)

Stretching or compressing cells can change which of their genes are switched on or off. But researchers have been unclear as to whether this occurs because signals are sent from the squashed cell membrane — or whether the nuclear envelope, which surrounds nuclear DNA, transmits the mechanical force.

To investigate this, Kris Noel Dahl *et al.* placed the nuclei of frog (*Xenopus*) egg cells in a solution and added polymer to control by osmosis how much the nucleus swelled. They found that the nuclear membrane can expand to almost twice its regular size.

The researchers also sucked up the membrane into a tiny pipette to test its strength, and concluded that it is stiffer and more resilient than the cell membrane, somewhat like a rubber glove. But the nucleus could not be made to compress down below its normal volume.

The authors conclude that the nuclear envelope offers a degree of protection for the DNA inside, while being flexible enough to transmit some forces when the cell's shape changes. They plan to examine whether stretching the nucleus and its envelope causes changes in gene activity, and how this is affected by mutations in certain nuclear-envelope proteins — for example lamins, which have been linked to human disorders such as premature ageing diseases.

Helen Pearson

Physiology

Stiffness and striation

J. Cell Biol. **166**, 877–887 (2004)

Adam J. Engler *et al.* have found that muscle fibres might be sensitive to the stiffness of their environment — a discovery that could affect future treatments of skeletal-muscle defects.

Immature muscle cells can be grown on surfaces of varying flexibility, fusing together to form muscle fibres. But in order to function properly, thick and thin filamentous proteins inside the fibres have to line up in a particular way, in a process called striation. Engler *et al.* show that this can only occur when the growth substrate is of medium stiffness, similar to that of normal muscle. On a stiff substrate such as glass, or a soft substrate such as a weak gel, striation is limited.

The authors also find that muscle samples from mice with a form of muscular dystrophy are stiffer than those from normal animals. The less flexible environment may limit striation, perhaps explaining why

Genomics

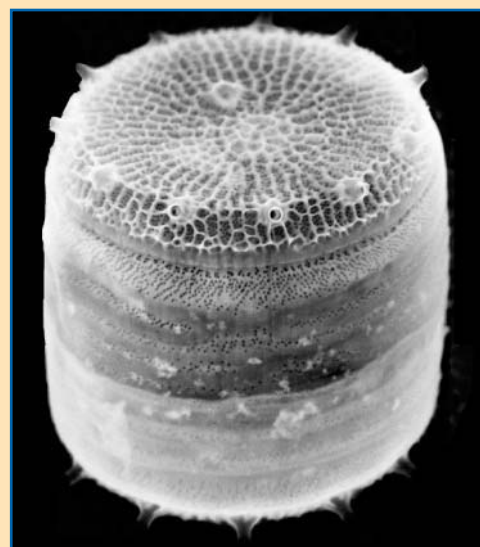
Lifestyle secrets of a diatom

Science **306**, 79–86 (2004)

Researchers have sequenced the genome of the marine diatom *Thalassiosira pseudonana*. Lurking within its 24 chromosomes and various assorted other DNA segments are genes that are involved in fashioning its ornate silica cell walls (pictured), utilizing iron and nitrogenous compounds, making fatty acids, and processing a complete urea cycle.

Diatoms are single-celled organisms found in both fresh water and the oceans. They are collectively responsible for 20% of the carbon fixed by photosynthesis throughout the globe — a contribution comparable to that of the world's rainforests. As E. Virginia Armbrust and colleagues now show, their diverse watery lifestyles are helped by their wide genetic repertoire.

The authors used the 'shotgun' whole-genome sequencing method to draw up a draft of *T. pseudonana*'s 34 million base pairs of nuclear DNA, as well as its 129,000-base-pair plastid genome and the 44,000 letters of its mitochondrial



code. The fact that it possesses genes for processing silica is no surprise — it has to build its intricate cell wall somehow. But the presence of so many different metabolic pathways, and the fact that half of the genes discovered have unknown functions, is testament to how little we understand about diatoms' eclectic metabolism.

Michael Hopkin

injured muscles in such mice are not repaired.

This may be bad news for those wishing to use stem-cell transplants to treat patients with muscular dystrophy. The stiff environment might prevent the cells from turning into contractile muscle, the authors warn.

Helen Pilcher

Particle physics

Rare prize

Phys. Rev. Lett. **93**, 131802 (2004)

Within the intricate filigree of the standard model, a *B* meson can decay into other particles in a variety of ways. The details of its decay patterns reveal 'CP violation' — an effect that shatters the delicate balance between matter and antimatter.

J. Dragic *et al.* report the first evidence of the decay of a neutral *B* meson (B^0) into two neutral particles, ρ^0 and π^0 . Using the Belle detector at the KEKB accelerator in Japan, the authors have collected more than 10^8 pairs of *B* mesons, produced in collisions between electrons and their antimatter counterparts, positrons. From this huge data set, Dragic *et al.* have pinpointed a $B^0 \rightarrow \rho^0 \pi^0$ signal representing just 15.1 ± 4.8 events.

This is a rare decay indeed; other experiments had succeeded only in setting limits on the probability of its occurrence. But its value comes in helping to pin down a vital parameter of the CP-violation process. A larger signal is needed to achieve that, but there are more data to come from this and competing experiments.

Allison Wright

Photoelectrochemistry

Molecular trees split water

J. Am. Chem. Soc. **126**, 12084–12089 (2004)

Polymers that sprout bushy branches all along their molecular chains prove to be efficient at channelling light energy into the production of hydrogen from water. Such photocatalytic water-splitting could one day fuel a hydrogen economy, if it can be done cheaply enough. Dong-Lin Jiang *et al.* (of the ERATO nanospace project) have used these 'dendrimeric' polymers in a photochemical process that relays electrons onto colloidal particles of platinum, where they are used to reduce water and liberate hydrogen gas.

The electrons are ferried to the platinum particles by a dye molecule called methyl viologen, in the form of a positively charged free radical. The polymer is the initial electron donor: capture of a photon of sunlight kicks an electron loose. Conjugated polymers could play this role in principle, but they have poor solubility and suffer from self-quenching — the liberated electrons tend to fall back into the 'holes' they leave behind. Wrapping the polymer chains in negatively charged, dendrimeric side-branches not only renders them soluble but also suppresses self-quenching. The efficiency of this system (measured by the fraction of H_2 molecules produced for each photon absorbed) is, at 13%, an order of magnitude better than obtained with previous dye-based photoreduction systems.

Philip Ball

Feather-pecking and victim pigmentation

A genetic factor that encourages this form of farmyard bullying has been identified.

Feather-pecking in domestic birds is associated with cannibalism and severe welfare problems¹. It is a dramatic example of a spiteful behaviour in which the victim's fitness is reduced for no immediate direct benefit to the perpetrator² and its evolution is unexplained. Here we show that the plumage pigmentation of a chicken may predispose it to become a victim: birds suffer more drastic feather-pecking when the colour of their plumage is due to the expression of a wild recessive allele at *PMEL17*, a gene that controls plumage melanization³, and when these birds are relatively common in a flock. These findings, obtained using an intercross between a domestic fowl and its wild ancestor, have implications for the welfare of domestic species and offer insight into the genetic changes associated with the evolution of feather-pecking during the early stages of domestication.

To investigate whether there could be a genetic basis for a chicken becoming a target of feather-pecking, we analysed the variance in feather damage due to feather-pecking suffered by individual birds (Fig. 1a; for methods, see supplementary information). These birds were from the second generation of a large-scale intercross between a line of white leghorn domestic fowl (*Gallus gallus domesticus*) and its wild ancestor, the red junglefowl *G. gallus*. We found that feather-pecking damage had a highly significant quantitative trait locus (QTL; residual variance explained, 14.9%) that coincided perfectly with the dominant white locus in the chicken genome⁴ (31 compared with 32 centimorgans; Fig. 1b).

The domestic fowl's dominant white allele (*I*) at this locus inhibits feather pigmentation, whereas birds homozygous for the wild-type allele (*i*) show a variety of pigmented (non-white) plumage. We found that these pigmented birds were significantly more vulnerable to feather-pecking than the white birds; a significant dominance effect revealed that heterozygous birds (*I/i*) that were white, or white with some minor pigmentation, were almost as well protected from feather-pecking as the *I/I* homozygotes (Fig. 1b, legend).

We showed previously that the dominant white colour is caused by a 9-base-pair insertion in exon 10 of the *PMEL17* gene, which encodes a melanocyte-specific protein that is essential for the normal development of eumelanosomes³. The fact that the QTL effect on feather-pecking and the *PMEL17* effect on pigmentation show the same mode of inheritance (the allele from the domestic

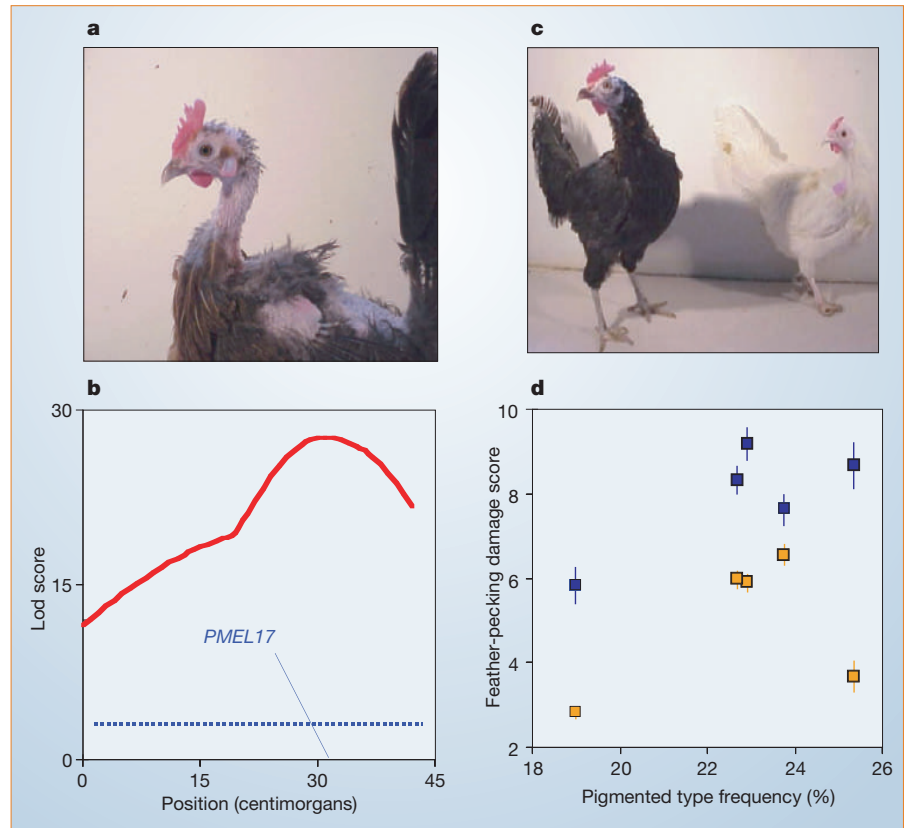


Figure 1 Chicken vulnerability to feather-pecking. **a**, Bird with feather damage due to feather-pecking by other birds. **b**, Lod scores for quantitative trait locus (QTL) on chromosome E22C19W28 associated with feather damage. The identified QTL coincides with the plumage-colour gene *PMEL17* and is significant at the genome-wide level (lod score, 26.3; $P < 0.0001$; dotted line: $P = 0.05$). Additive (1.67 ± 0.18) and dominance (-1.25 ± 0.29) effects ($\pm$ s.e.) show that feather damage is 2.92 and 3.34 times higher in *i/i* than in *I/I* and in *I/i* birds, respectively: the tendency for *I/i* to suffer more feather damage than *I/I* birds was not significant. **c**, *PMEL17* plumage phenotypes: domestic, white type (*I/I*, right) and wild, pigmented type (*i/i*, left). **d**, Mean feather damage in pigmented (*i/i*; blue) and white (*I/I* and *I/i*; orange) birds in relation to the frequency of pigmented birds in a cohort; bars indicate standard error.

chicken is almost completely dominant) indicates that *PMEL17* is likely to be the causative gene for this QTL. We conclude that the victim's phenotype is critical for stimulating this spiteful behaviour.

Social transmission⁵ may contribute to the spread of feather-pecking when genetically predisposed victims are relatively common. We find an environmental effect that contributes to feather-pecking damage and is associated with the frequency of the two *PMEL17* plumage phenotypes (pigmented and white; Fig. 1c) in a cohort: pigmented birds are more vulnerable to feather-pecking than white birds when they are relatively common (Fig. 1d). Vulnerability to feather-pecking may therefore also be influenced by the relative frequency of the two *PMEL17* phenotypes, directly and through interaction with a bird's phenotype (cohort: $\chi^2_{4,720} = 122.52$, $P < 0.0001$; phenotype: $\chi^2_{1,720} = 112.79$, $P < 0.0001$; cohort $\times$ phenotype: $\chi^2_{4,720} = 11.91$, $P = 0.018$; Fig. 1d). In

addition, the light-coloured wood-shavings that are used as floor litter may be more conspicuous when deposited on pigmented rather than on white birds⁶ and so may act as a pecking stimulus.

The propensity to peck feathers is independent of the assailant's plumage genotype in the intercross studied here⁷. Some genetic variation associated with becoming a victim of feather-pecking has been found among other domestic lines^{8–10}. The propensity to perpetrate or to become a victim of feather-pecking therefore seems to be genetically independent.

We have discovered a strong genetic effect that is mediated by plumage phenotype. Our findings illustrate the importance of a victim's phenotype in the evolution of spiteful behaviours such as feather-pecking and have crucial implications for the welfare of domestic species. Key issues still to be addressed are the extent to which differently pigmented plumage phenotypes (for example, black or

barred) predispose a bird to feather-pecking and the way in which this is influenced by the social environment through frequency-dependent effects.

Linda Keeling*, Leif Andersson†‡, Karin E. Schütz*§, Susanne Kerje†||, Robert Fredriksson‡, Örjan Carlborg‡¶, Charles K. Cornwallis#, Tommaso Pizzarilli*, Per Jensen||

*Department of Animal Environment and Health, Swedish University of Agricultural Sciences, PO Box 234, 53223 Skara, Sweden

†Department of Medical Biochemistry and Microbiology, Uppsala University, Box 597, 75124 Uppsala, Sweden

‡Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, 75124 Uppsala, Sweden

§AgResearch, Animal Behaviour & Welfare, Private Bag 3123, Hamilton, New Zealand

||Department of Biology, IFM, Linköping University, 58183 Linköping, Sweden

e-mail: per.jensen@ifm.liu.se

¶Roslin Institute, Roslin, Midlothian EH25 9PS, UK

#Department of Animal & Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

☆Evolution & Ecology Group, School of Biology, University of Leeds, Leeds LS2 9JT, UK

1. Blokhuys, H. J. & Wiepkema, P. R. *Vet. Quart.* **20**, 6–9 (1998).
2. Johnstone, R. A. & Bshary, R. *Proc. R. Soc. Lond. B* **271**, 1917–1922 (2004).
3. Kerje, S. *et al. Genetics* (in the press).
4. Smyth, J. R. in *Poultry Breeding and Genetics* (ed. Crawford, R. D.) 109–167 (Elsevier Science, New York, 1996).
5. McArdie, T. M. & Keeling, L. J. *Appl. Anim. Behav. Sci.* **75**, 147–159 (2002).
6. Savory, C. J. & Mann, J. S. *Brit. Poul. Sci.* **40**, 565–572 (1999).
7. Jensen, P. *et al. ISAE 2003 Meeting*, June 24–28, 68 (2003).
8. Kjaer, J. B. & Sørensen, P. *Brit. Poul. Sci.* **38**, 333–341 (1997).
9. Buithenhuys, A. J. *et al. Poul. Sci.* **82**, 1661–1667 (2003).
10. Buithenhuys, A. J. *et al. Poul. Sci.* **82**, 1215–1222 (2003).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Pair bonds

Arrival synchrony in migratory birds

Synchronous arrival of pairs of migratory birds at their breeding grounds is important for maintaining pair bonds and is achieved by pairs that remain together all year round. Here we show that arrival is also synchronized in paired individuals of a migratory shorebird, the black-tailed godwit (*Limosa limosa islandica*), even though they winter hundreds of kilometres apart and do not migrate together. The mechanisms required to achieve this synchrony and prevent 'divorce' illustrate the complexity of migratory systems.

Long-lived migratory birds generally show high degrees of mate fidelity, and divorce is often followed by a decrease in reproductive success¹. Synchrony in timing of arrival on the breeding grounds is thought to be crucial for retaining a mate from the

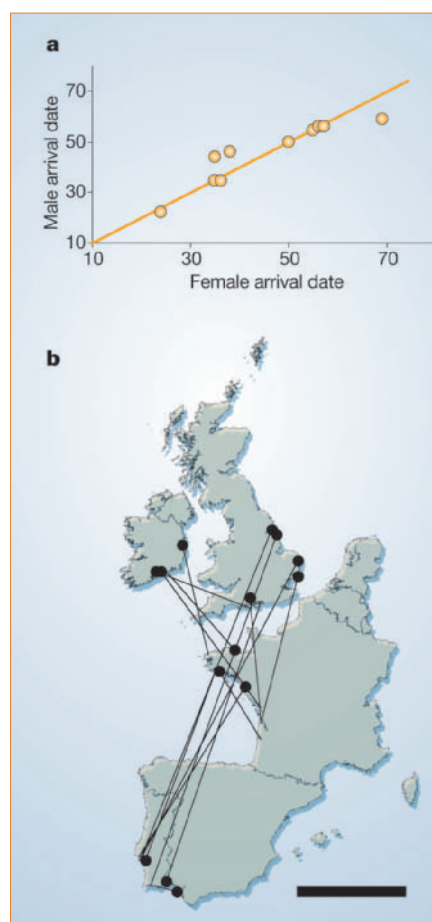


Figure 1 Paired black-tailed godwits arrive at their breeding sites synchronously but winter separately. **a**, The arrival dates (days from 31 March) of paired males and females were highly correlated ($r=0.97$, $P<0.001$); the regression line indicates identical arrival times for males and females. For each pair, data are for one year only. **b**, Wintering locations of pairs of godwits. Each line links a pair and the circle indicates the female's location. Locations of 12 of the 14 pairs are known in the same winter. Godwits show a very high degree of between-year philopatry in winter.

previous year and avoiding a costly divorce^{2,3}. In a few species¹, pairs both migrate and winter together, so synchronous arrival in spring is inevitable. In many others, one sex departs from the breeding grounds ahead of the other and so the pair migrate separately⁴; however, their arrival could still be synchronous if they wintered together or were reunited during the spring migration.

The Icelandic black-tailed godwit winters between Britain and Iberia and breeds almost exclusively in Iceland⁵. About 1.5% of the population is currently individually colour-marked⁵, and the winter destinations of 55% of these birds have been identified using a large network of volunteer observers. To investigate the arrival patterns of pairs of godwits, we located individuals that had been colour-marked as breeding birds in previous years by twice-weekly searches of 14 study plots on the breeding grounds in southern Iceland, in 2002 and 2003.

Breeding godwits arrived over a one-month period between mid-April and the

middle of May. Previously paired males and females arrived within 3.1 (± 1.3 s.e., $n=10$) days of one another (Fig. 1a). These remarkably synchronous arrival times are not achieved through pairs wintering in the same area and therefore departing together: paired male and female godwits winter, on average, 955 km apart (± 165 s.e., $n=14$; range, 49–1,946 km; Fig. 1b).

Neither do the pairs meet during migration: during 1999–2004, we studied migratory flocks upon their arrival in Iceland (before they moved on to the breeding grounds)^{5,6} and never encountered paired birds in the same flocks, despite locating 15 marked individuals whose mates were known. Arrival synchrony seems to be related to mate retention, however, as the only divorces occurred in two of the three pairs that arrived more than eight days apart (Fig. 1a).

How is this degree of synchrony maintained between pairs when they winter so far apart and the environmental conditions for migration are likely to differ locally? It is possible that pairs of birds may winter in areas of similar quality (despite their geographic separation) and so be in a similar condition to arrive at specific times in spring; or they may share some genetic or physiological determinant of timing of migration; or they may independently synchronize their arrival to the optimal time for each specific breeding location (for example, to exploit peaks in resource abundance). As individuals often use a series of passage sites during spring migration, they may refine these timings as they approach their breeding grounds. Identifying which of these mechanisms is operating is likely to be key to understanding how synchrony is achieved and divorce avoided in migratory species.

T. G. Gunnarsson*, J. A. Gill*†, T. Sigurbjörnsson‡, W. J. Sutherland*

*Centre for Ecology Evolution and Conservation, School of Biological Sciences, University of East Anglia, and †Tyndall Centre for Climate Change Research, Norwich NR4 7TJ, UK

e-mail: t.gunnarsson@uea.ac.uk

‡Langhús, Fljót 570, Iceland

1. Ens, B. J., Choudhury, S. & Black, J. M. in *Partnerships in Birds: the Study of Monogamy* (ed. Black, J. M.) 344–401 (Oxford Univ. Press, Oxford, 1996).
2. Coulson, J. C. & Thomas, C. S. *Proc. Int. Ornithol. Congr.* **17**, 823–833 (1980).
3. Handel, C. M. & Gill, R. E. *Anim. Behav.* **60**, 471–481 (2000).
4. Reynolds, J. D. & Székely, T. *Behav. Ecol.* **8**, 126–134 (1997).
5. Gunnarsson, T. G. *et al. Bird Study* (in the press).
6. Gill, J. A. *et al. Nature* **412**, 436–438 (2001).

Competing financial interests: declared none.

brieF communications is intended to provide a forum for short, topical reports of general scientific interest and for online technical discussion of recently published material of particular interest to non-specialist readers (**brieF communications arising**). Priority will be given to contributions that have fewer than 500 words, 10 references and one figure. Detailed guidelines are available on Nature's website.

Surface mechanics mediate pattern formation in the developing retina

Takashi Hayashi^{1,2} & Richard W. Carthew¹

¹Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208, USA

²Department of Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan

Pattern formation of biological structures involves organizing different types of cells into a spatial configuration. In this study, we investigate the physical basis of biological patterning of the *Drosophila* retina *in vivo*. We demonstrate that E- and N-cadherins mediate apical adhesion between retina epithelial cells. Differential expression of N-cadherin within a sub-group of retinal cells (cone cells) causes them to form an overall shape that minimizes their surface contact with surrounding cells. The cells within this group, in both normal and experimentally manipulated conditions, pack together in the same way as soap bubbles do. The shaping of the cone cell group and packing of its components precisely imitate the physical tendency for surfaces to be minimized. Thus, simple patterned expression of N-cadherin results in a complex spatial pattern of cells owing to cellular surface mechanics.

Pattern formation is the process by which cells are spatially arranged into elaborate biological structures. Whereas significant advances have been made in understanding how different cell types become determined, less is understood about how cells are configured into the characteristic shapes found in many organs. Signal transduction networks under genetic control are known to regulate cell shape and form through the cytoskeleton^{1,2}. On the other hand, sporadic descriptive analysis over the past century has suggested that cells can be configured by mechanisms following physical principles. The notion that cells organize themselves within structures that minimize their total surface area is a recurring one.

The similarities between the patterns adopted by simple cell aggregates and soap bubbles have long been remarked upon³. A set of rules governing the behaviour of soap bubbles was first developed by Plateau⁴. He found that in stable systems of bubbles, junctions of more than three bubbles at a point were never observed. He observed that three and only three bubbles meet at any junction within a two-dimensional array of bubbles. The mathematical and physical justifications for this phenomenon are based on minimizing surface area for a given group of bubbles⁵. Plateau also observed that the meeting point of any three bubbles was symmetrical. The surface free energy of each of the three contact surfaces is equal to the others, meaning that the three surfaces meet at an intersection point, forming angles of 120° (Fig. 1a). Because the bubble membranes (soap films) are identical, the interactions between neighbouring bubbles are identical and little complexity can be generated by the system. However, many simple tissues and multicellular aggregates in nature that closely resemble the packing pattern seen with soap bubbles have been documented³ (Fig. 1b). This correlation suggested that cells pack together in configurations that minimize their overall surface area.

A different manifestation of surface minimization is seen when mixtures of dissociated embryonic cells aggregate *in vitro*. As first pointed out by Steinberg⁶, cells aggregate in ways that closely imitate the mixing of immiscible liquids. When two immiscible liquids are mixed together, the liquid with the greater surface energy will round up into a circle or sphere surrounded by the other. This shape minimizes the liquid's surface area. Likewise, dissociated cells migrate and aggregate into an overall spherical shape if they are able to adhere to one another; this process correlates with the aggregate's total surface energy^{7–9}. Thus, a group of cells will adopt an overall shape that minimizes the group's surface area.

A wide variety of biological structures might be patterned following minimization principles. Cells in epithelial sheets and

early embryos join in three-way junctions, as do lobules in a liver. Epidermal scales and hair follicles are organized as circular structures, and cell groups within glands frequently adopt circular shapes. Despite this circumstantial evidence, there is limited experimental data that surface mechanics influences patterning *in vivo*. Moreover, many complex biological patterns emerge in epithelia

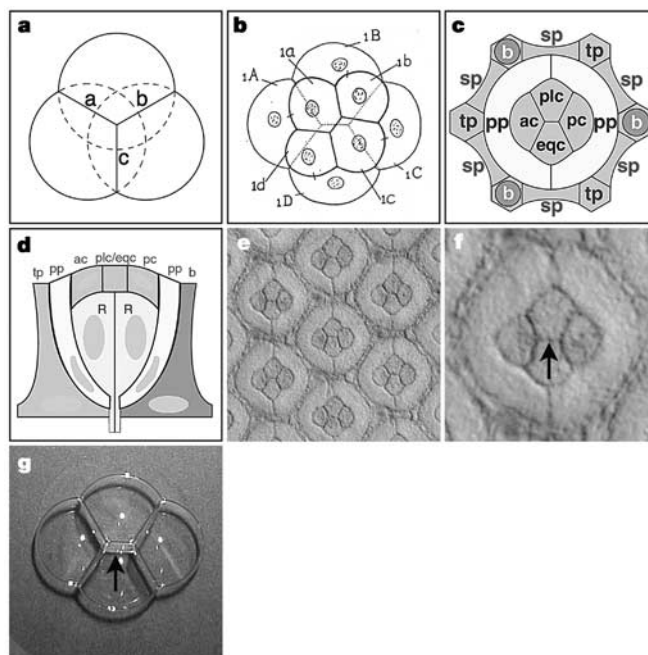


Figure 1 Pattern formation in different tissues. **a**, The configuration of three soap bubbles. The interfaces (a,b,c) are oriented at 120°-angles from each other. **b**, Packing symmetry in an eight-cell mollusc embryo⁴⁰. **c, d**, Schematic *Drosophila* ommatidium at 35% of pupal life. Cross-section view at the level of the adherens junction (**c**), and the side view is equatorial to the midplane (**d**). Cone cells (eqc, equatorial; pc, posterior; plc, polar; ac, anterior) are surrounded by two primary pigment cells (pp) plus secondary pigment (sp), tertiary pigment (tp) cells and bristles (b). The cone cells sit over a cluster of photoreceptor cells (R). **e**, A retina stained with cobalt sulphide. **f**, An ommatidium with an arrow marking the junctional interface between cone cells. **g**, A cluster of four soap bubbles, with an arrow marking the junctional interface.

where cells are physically constrained. Epithelial cells are anchored at the apical and basal surfaces, and they are not free to move with respect to their neighbours. Thus, it is unclear if pattern formation in epithelia follows constraints imposed by their surface properties. Even so, mathematical modelling has attempted to describe epithelial pattern formation using Plateau's rules and minimization principles¹⁰.

In this study, we investigate the physical basis of biological patterning in the *Drosophila* retinal epithelium. We demonstrate that E- and N-cadherins mediate apical adhesion between cells in the epithelium. Differential expression of N-cadherin within a subgroup of retinal cells (cone cells) results in the cone cell group adopting an ellipse/lozenge form that is surrounded by cells not expressing N-cadherin. Moreover, in both normal and experimentally manipulated conditions, cone cells pack together within the form in the same way as soap bubbles. Thus, pattern is determined by mechanisms that minimize overall surface area.

Pattern formation in the *Drosophila* retina

The *Drosophila* retina is an epithelium composed of a hexagonal array of facets called ommatidia. Each ommatidium consists of twenty cells arranged in a precise pattern, including eight photo-

receptor neurons and twelve accessory cells¹¹. Four cone cells form a plate that acts as both the floor of the simple lens and the roof of the underlying chamber of photoreceptors (Fig. 1c, d). Two primary pigment cells form the walls of the lens, shaping and optically insulating it. Other pigment cells and bristle cells form the ommatidium border, lending it a distinctly hexagonal shape. This pattern is evident near the apical surface of the epithelium, coincident with the adherens junction (Fig. 1e). Morphogenesis of the pattern occurs primarily during pupation^{12,13}, and it does not involve cell migration or sorting¹¹. The primary pigment cells enlarge and begin to surround the cone cells, displacing other cells from contact. Eventually they ensheath the cone cells and make contact with each other. Meanwhile, the polar and equatorial cone cells expand between the anterior and posterior cone cells and contact each other. The net result is a transparent plate that functions as an aperture for focused light to transmit from lens to photoreceptors. This pattern is of fundamental importance to the compound eye as it is strongly conserved from Crustacea to Insects¹⁴.

Soap bubble modelling of cone cell patterns

The configuration of cone cells superficially obeys Plateau's rules. The four cells do not intersect at a single point. Rather, two points of

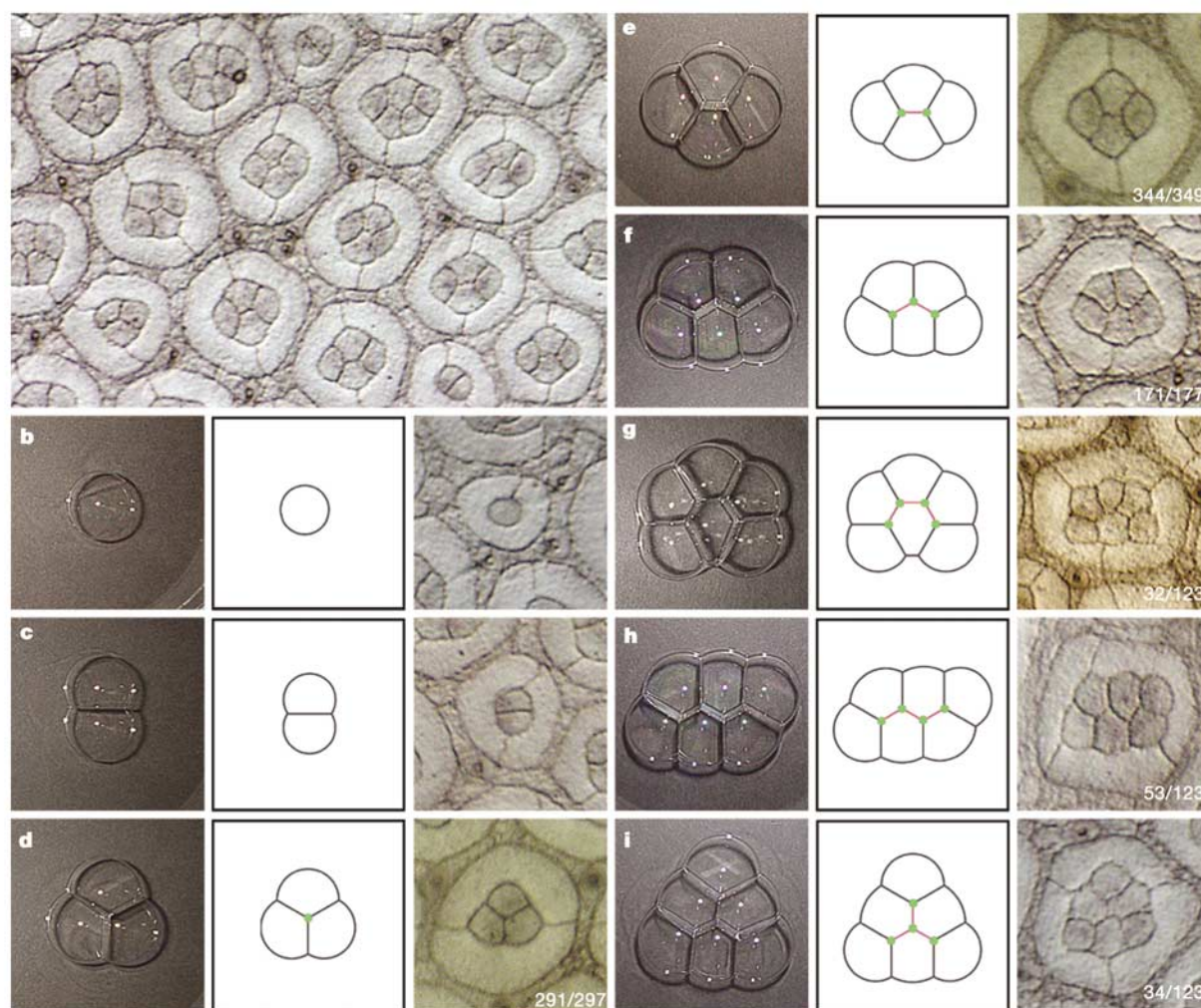


Figure 2 Configuration of cone cells precisely correlates with soap bubble configurations. **a**, Retina from a *Roi/+* fly at 35% of pupal life. Ommatidia contain variable numbers of cone cells. **b–i**, Left panels: stable soap bubble aggregates. Middle panels: schematics of bubble patterns. Green dots represent three-way intersection points; red lines represent junctional interfaces; black lines represent other membrane interfaces. Right panels:

typical *Roi/+* cone cell groups. Inside the panels are the indicated numbers of *Roi/+* groups with that particular configuration compared with the total number in each group. Aggregates of one (**b**), two (**c**), three (**d**), four (**e**), five (**f**) and six (**g–i**) cone cells. For ommatidia with one or two cone cells, 100% exhibited the configurations shown in (**b**) and (**c**).

intersection each join three cells together (Fig. 1f). The two intersection points are connected by an interface between the polar and equatorial cells. This interface is unique from the other cone cell interfaces because it connects the intersection points. We refer to this type of interface as a junctional interface. The intersection points display threefold symmetry; each interface is oriented approximately 120° from other interfaces. Thus, the configuration of cone cells strongly resembles the stable configuration of four soap bubbles (Fig. 1g).

To determine whether cone cells intrinsically configure themselves into structures that minimize their overall surface area (like soap bubbles), we examined ommatidia with different numbers of cone cells and compared the cone cell patterns with the patterns adopted by corresponding numbers of soap bubbles (see Methods). Plateau⁴ observed that soap bubbles aggregate into stable and predictable configurations depending on the number of bubbles in each group. In an aggregate of more than two bubbles, the numbers of intersection points $N(I)$ and junctional interfaces $N(J)$ are: $N(I) = N - 2$ and $N(J) = N - 3$ (where N is the number of bubbles in the aggregate). For example, an aggregate of five bubbles has three intersection points and two junctional interfaces. In addition, all intersection points exhibit threefold symmetry, independent of the number of bubbles. Thus, junctional interfaces, like other interfaces, tend to incline to one another at angles of 120° . Together, these properties restrict the number of possible stable configurations that soap bubbles adopt. There is only one stable configuration for aggregates of five or fewer bubbles, and there are three stable configurations for six bubbles (Fig. 2).

To determine whether cone cells also organize into configurations governed by surface minimization forces, we examined the ommatidia of *Rough eye* (*Roi*) mutants. *Roi* ommatidia contain variable numbers of cone cells (Fig. 2a) owing to a dominant mutation in the bHLH gene *amos*^{15,16}. We found that *Roi* cone cell clusters showed the same properties as soap bubbles in the number of intersection points, junctional interfaces and their inclinations (Fig. 2b–i). In ommatidia containing one cone cell, the cell had a radial interface with its surroundings, whereas clusters of two cells had a single straight interface between the cells, with their remaining surfaces curved (Fig. 2b, c). This precisely matches the forms shown by one or two soap bubbles. Similarly, for cone cell clusters ranging in number from three to five, a single configuration that exactly matches the corresponding bubble aggregate was primarily observed (Fig. 2d–f). For six-cell clusters, three configurations were primarily observed, and these exactly matched the three stable configurations seen with aggregates of six soap bubbles (Fig. 2g–i). This extensive correlation between cone cell and soap bubble configurations strongly indicates that the configurations are controlled by the tendency for cone cell groups to minimize their overall surface area.

DE- and DN-cadherins are essential for configuring cone cells

Soap bubbles aggregate by fusing their film surfaces along bubble–bubble interfaces. Membrane fusion does not occur between cone cells¹². How then do cone cells aggregate together? Cadherins are thought to play critical roles in allowing cells to connect together into coherent groups. Classical cadherins are transmembrane proteins localized to the adherens junction that mediate cell–cell adhesion by homophilic interactions^{17,18}. Strong adhesion requires both the extracellular cadherin domain and cadherin's ability to interact with the cytoskeleton. The cytoplasmic domain of cadherin associates with β -catenin, which binds to α -catenin, which connects to the actin cytoskeleton. In this manner, cytoskeleton form and shape is dependent upon cadherin localization. To determine whether the three classical *Drosophila* cadherins^{19–21} are involved in retinal cell patterning, we first examined their expression. CadN2 is not expressed in the developing pupal retina (data not shown). However, DE-cadherin is expressed in all retinal cells and is localized

to the adherens junction of all cell–cell interfaces (Fig. 3a–c). DN-cadherin is also expressed in the retina but is restricted from all pigment cells (Fig. 3d–f). Within cone cells, it is localized to the adherens junction of cone–cone interfaces. These localization patterns are consistent with the homophilic interactions typical of cadherins.

We next examined the patterning of retinal cells mutant for *DE-cadherin*. Small clones of cells mutant for *DE-cadherin* were generated shortly after the onset of retinal differentiation. The

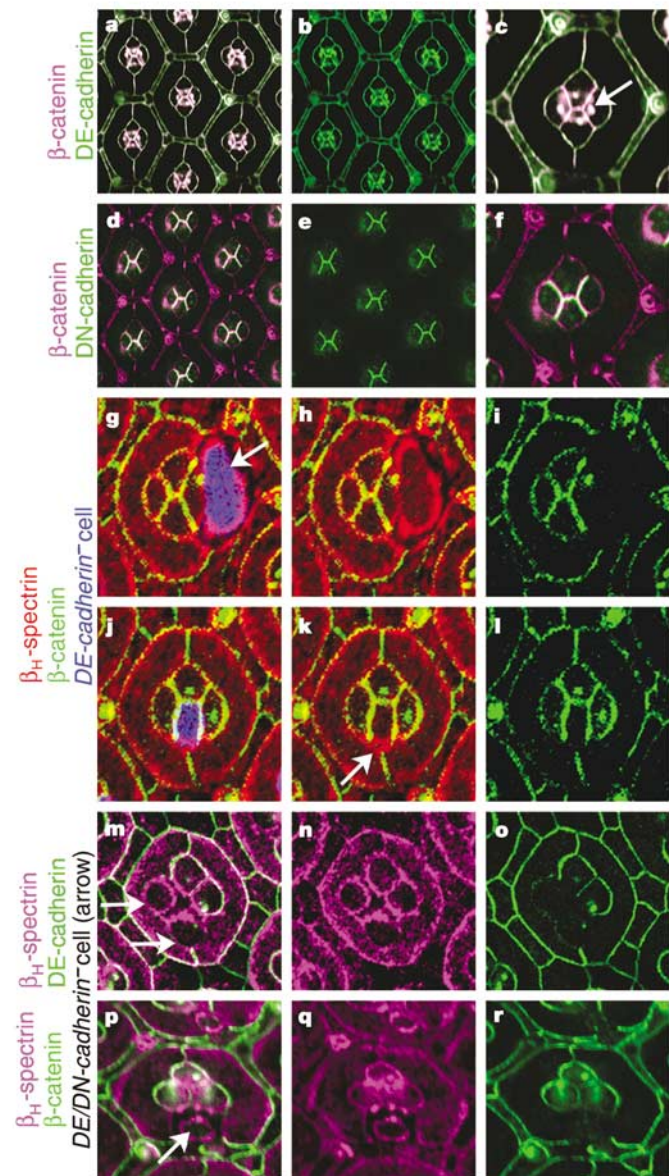


Figure 3 DE- and DN-cadherins are required for apical adhesion of retinal cells. **a–c**, Localization of DE-cadherin (green) and β -catenin (purple) proteins. Co-localized proteins appear white. Arrow in magnified ommatidium (**c**) shows projections from underlying photoreceptors. **d–f**, Localization of DN-cadherin (green) and β -catenin (purple) proteins. **g–i**, *DE-cadherin* mutant cells (*DE-cadherin*[−]) marked with GFP (blue), β -catenin (green) and β H-spectrin (red) outline membranes. A mutant primary pigment cell has lost contact with its neighbours (**g–i**). A mutant cone cell has lost contact (arrow) with primary pigment neighbours (**j–l**). **m–r**, Mosaic ommatidia with cone cells doubly-mutant for *DE-* and *DN-cadherin* (arrows), stained for β H-spectrin (purple) and DE-cadherin (green). A mutant cone cell has lost complete apical contact with its neighbours (**p–r**).

membranes of mutant primary pigment cells did not adhere to neighbouring cells (Fig. 3g–i). Lack of β -catenin in the mutant cells indicated an absence of a functional adherens junction. Mutant cone cells failed to adhere to primary pigment cells but still adhered to neighbouring cone cells (Fig. 3j–l). We conclude that DE-cadherin is required in both cone and primary pigment cells for adhesion. Because DN-cadherin is expressed in cone cells along cone–cone interfaces, we proposed that this molecule was also involved in cone cell adhesion. To test this, we induced clones of *DE-cadherin* mutant cells that were also mutant for *DN-cadherin*. The *DN-cadherin* allele is a partial loss-of-function mutant, allowing us to recover flies at the appropriate stage. Doubly mutant cone cells failed to adhere to their cone and primary pigment cell neighbours (Fig. 3m–r). Some mutant cells lost total apical contact with other cells (Fig. 3p–r). It is unclear at present what occupies the space between mutant cells and their neighbours before tissue fixation and staining. Generally, the mutant cone cells exhibited a radial interface with their surroundings. In such mosaic ommatidia, the remaining wild-type cone cells packed into configurations typical for soap bubbles of similar number (compare Fig. 3p to Fig. 2d). Thus, DE- and DN-cadherins are essential for cone cells to connect and pack into their canonical configuration.

DN-cadherin regulates the shape of cone cell groups

To understand the specific function of *DN-cadherin*, we examined clones of null mutant cells (Fig. 4a, b). Loss of *DN-cadherin* had no detectable effect on primary pigment cells. When all cone cells in an ommatidium were mutant for *DN-cadherin*, cells maintained adherens contact with their neighbours at the centre of the ommatidium.

Occasionally, an individual cone cell was seen separated from its group (Fig. 4h). However, the predominant phenotype observed was a change in the shape of a cone cell group from its normal lozenge/ellipse-like form into a shape that resembled a cruciform (Fig. 4a, c, g). All mutant ommatidia exhibited this shape change, or more rarely, cone group separation ($N > 500$ ommatidia examined). The alteration to a cruciform shape was accompanied by two key changes in interfacial geometry (Fig. 4j–l). First, the length of cone–cone interfaces was 36% shorter in *DN-cadherin* mutants than in wild type ($P < 0.001$, Mann–Whitney; Fig. 4j, k). The length of cone–primary pigment interfaces was correspondingly longer. Second, cone–primary pigment interfaces were oriented to each other by an average of 154° (s.d. = 11°) in wild-type ommatidia, whereas they were oriented to each other by 108° (s.d. = 11°) in *DN-cadherin*[−] ommatidia (Fig. 4j, k).

We further examined ommatidia that were composed of one or two mutant cone cells. In such mosaic ommatidia, cells behaved with strict autonomy according to their phenotypes (Fig. 4d, e). A single mutant cone cell in an otherwise wild-type ommatidium adopted a mutant morphology. A mosaic group of two mutant cells paired with two wild-type cells adopted a chimaeric cruciform–ellipse configuration. These results indicate that DN-cadherin is required cell-autonomously and not as a signal to alter cell shape. However, DN-cadherin expressed in one cell was not sufficient to impart the cell with normal morphology unless one of its neighbours also expressed DN-cadherin (Fig. 4f), and is consistent with DN-cadherin acting through homophilic interactions between neighbouring cells. Thus, clonal analysis suggests that DN-cadherin regulates the shape of a cone cell group.

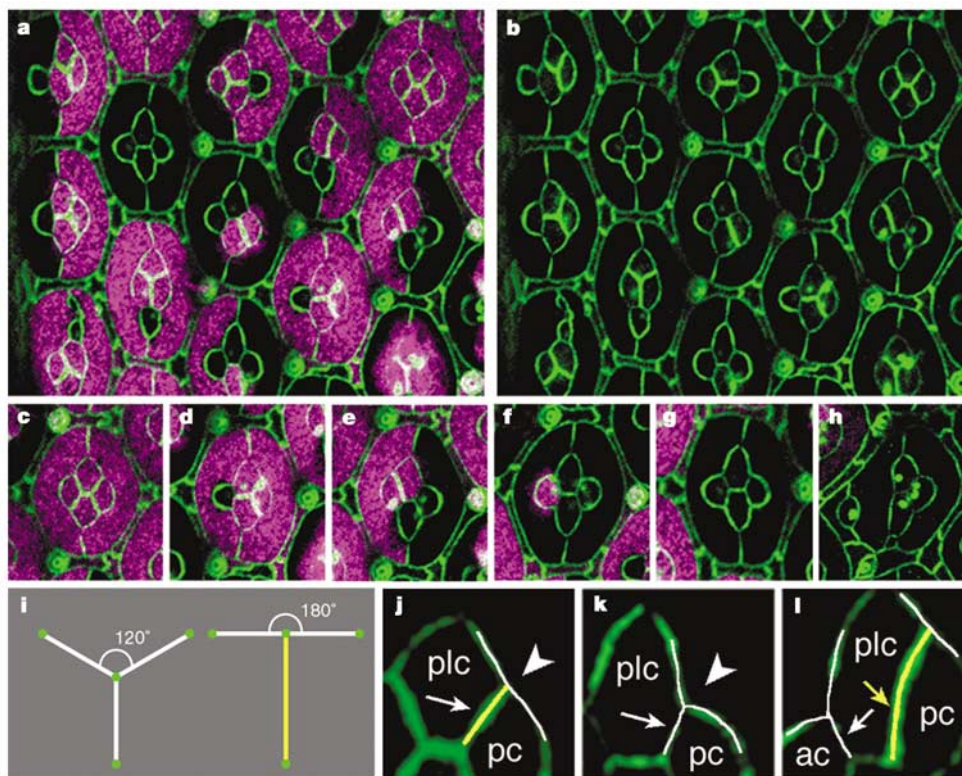


Figure 4 DN-cadherin is required for cone cell patterning. *DN-cadherin* mutant cells are marked by the absence of LacZ (purple) and visualized with β -catenin (green). **a, b**, A mosaic retina. **c–h**, Mosaic ommatidia, showing patterns when zero (**c**), one (**d**), two (**e**), three (**f**) and four (**g, h**) cone cells are mutant. **i**, Schematic of three interfaces that meet at a point. Left, adhesion is equivalent among the interfaces. Right, one interface (yellow) has greater adhesion than the other interfaces. **j, k**, Three-way interfaces between a primary

pigment, a polar- and posterior-cone cell are highlighted. Primary pigment interfaces meet (arrowhead) at an angle close to 180° when cells are wild type (**j**), and close to 120° when cells are mutant (**k**). The cone–cone interface (arrow) is shorter in the mutant. **l**, A mosaic ommatidium with an interface between wild-type cells (yellow), and an interface between a mutant and wild-type cell (white).

If differential expression of DN-cadherin is responsible for the cone cell group's shape, then providing DN-cadherin to surrounding primary pigment cells would alter the shape of the cone cell group. When DN-cadherin was misexpressed in both primary pigment cells surrounding a cone cell group, the cone cells adopted a cruciform structure that was indistinguishable from cruciforms seen in *DN-cadherin* mutants (Fig. 5a, h). Thus, cruciform formation correlates with uniform DN-cadherin expression. In contrast, overexpression of DE-cadherin did not generate cruciform-like structures (Fig. 5b), even though DE-cadherin levels were much higher in overexpressing cells (data not shown). The only phenotype we observed were junctional interfaces forming between anterior and posterior cone cells that overexpressed DE-cadherin. Because overexpression of DE-cadherin blocks Wnt signalling and leads to cell fate changes²², we determined whether cell-type markers were properly expressed in pigment and cone cells overproducing DN- or DE-cadherin. The expression of Cut and BarH1 proteins was normal in retinal cells (data not shown). Thus, differential expression of DN-cadherin mediates shaping of a cone cell group without altering its identity.

How is group shape affected by DN-cadherin? Cadherins promote cell–cell adhesion through their extracellular domains²³. Another mechanism by which DN-cadherin might potentially alter cone cell shape is through the actin cytoskeleton. To address whether DN-cadherin affects the shape of a cone cell group through its adhesive or cytoskeletal activities, we examined ommatidia in which a single primary pigment cell misexpressed DN-cadherin. In such ommatidia, there were major changes in cone cell position and shape (Fig. 5a, d–g). Cone cells preferentially contacted the primary pigment cell that misexpressed DN-cadherin, resulting in marked repositioning of the cone cells. Cone cells adopted several different shapes and attached themselves selectively to the cadherin-positive primary pigment cell. The positive primary pigment cells were altered into a variety of shapes, frequently engulfing one or more cone cells. On the basis of these data, we conclude that DN-cadherin drives greater contact between cells by a differential adhesion mechanism.

Regardless of whether DN-cadherin was misexpressed in one or both primary pigment cells, the DN-cadherin-positive cells (cone and primary pigment) coalesced into overall structures that were

lozenge-like or elliptical in shape (Fig. 5). The effect was completely penetrant for those ommatidia examined ($N \approx 100$). This observation argues that the shaping of cell groups is primarily determined by adhesion mediated by DN-cadherin to minimize the overall surface area of the group.

Discussion

In this study, we investigated the basic mechanisms of spatial patterning in an intact neuroepithelium. Cadherin expression is critical for this patterning at two levels. First, DE- and DN-cadherins are necessary for apical adhesion between cone cells. This permits cone cells to configure themselves in a manner that precisely matches the behaviour of surface films (that is, soap bubbles). Because soap bubbles pack together in configurations that minimize their overall surface area, this suggests that cone cell groups develop into configurations that behave similarly. A second way that cadherin functions is by shaping the cone cell group. A group of cells expressing DN-cadherin approximates an ellipse/lozenge form, surrounded by a group of cells lacking DN-cadherin. Analogous to a liquid film, this form is energetically favoured because the surface area of contact between the more adhesive group and surrounding cells approaches a minimum. Adoption of this form is accompanied by certain geometric characteristics of cone cell interfaces. Interfaces between cone and primary pigment cells meet at more obtuse angles, and interface length between neighbouring cone cells increases when DN-cadherin is localized along the interface (Fig. 4i–k).

The similarity between cone cell patterns and those formed by simple systems that minimize surface energies does not mean that the underlying physics of these two processes is the same. Inter-molecular attraction between detergent molecules provides surface free energy to bubbles. In an aggregate of bubbles, interfaces between neighbouring bubbles become single membranes, thus reducing the total surface area and consequently, surface free energy. In an aggregate of cells, adhesion molecules link cell membranes together, contributing to their surface free energy. When the area of adhesion spreads (increasing the cell–cell interface), this surface free energy is reduced. A link between interface length and adhesion is noticeably seen in *DE-cadherin* mutants, which exhibit both reduced cone–primary pigment adhesion and shorter cone–primary pigment interfaces (Fig. 3g, j).

Although a simple relation between surface mechanics and pattern is suggested, the system is probably more complex, on the basis of two observations. First, simply increasing cadherin levels in cells is not sufficient to organize shaping of cell groups, because overexpression of DN-cadherin but not DE-cadherin affected shaping (Fig. 5). Mammalian cadherins have also been found to differentially affect cell adhesion and sorting *in vitro*²⁴. Second, groups of DN-cadherin-positive cells do not always pack like unconstrained soap bubbles (Fig. 5). These groups might use other configuring mechanisms, or possibly additional constraints limit their ability to minimize surface area.

Our study demonstrates the interplay of genetic and physical mechanisms in pattern formation of epithelial tissues. Genetic controls determine the differential synthesis of cadherins in distinct sets of cells. Signalling pathways that are active in cells can further regulate the localization of intercellular adhesion proteins through adherens junction complexes^{2,25}. The regulated distribution of cadherins in turn allows cohesive groups of cells to minimize their surface area by forming overall ellipse/lozenge shapes. Cells within the group pack into configurations following Plateau's rules, achieving the lowest overall surface area. Our results demonstrate how the simple expression pattern of two adhesion proteins can render a complex pattern of cells in a developing biological structure. We suggest that similar relationships between gene expression and the physical properties of cells might contribute to the extraordinary diversity of patterns seen in biology. □

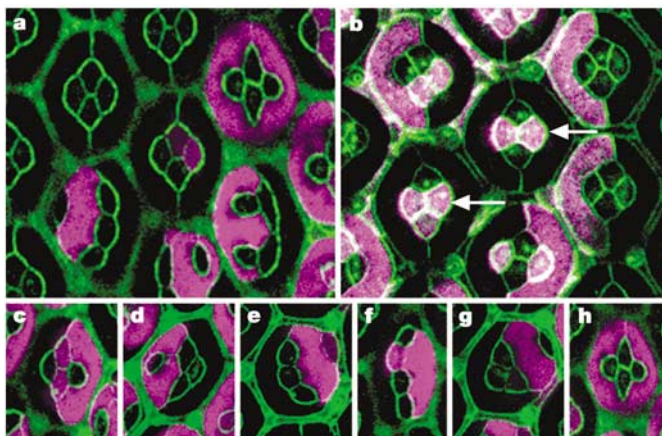


Figure 5 Misexpression of cadherins leads to patterning defects. **a**, Cells misexpressing *DN-cadherin* are marked with GFP (purple) and visualized with DE-cadherin (green). **b**, Cells misexpressing *DE-cadherin* are marked with GFP (purple) and visualized with β -catenin (green). Arrows mark ommatidia in which junctional interfaces are oriented between anterior and posterior cone cells rather than polar and equatorial cells. **c–g**, Examples of ommatidia in which one primary pigment cell misexpresses *DN-cadherin* (purple). **h**, A typical ommatidium in which both primary pigment cells misexpress *DN-cadherin* (purple).

Methods

Soap bubbles

A mixture of glycerol and dishwashing detergent in the ratio of 7:1 was used as the soap solution. Soap bubbles of equal volume were individually blown into aggregates on a polystyrene dish.

Histochemistry

Pupae were collected at the white-pupa stage and reared at 25 °C for 40 h (35% pupal life) before staining. Cobalt sulphide staining and antibody staining of pupal retinas were carried out as described²⁶. The antibodies used were as follows: rat anti-DE-cadherin¹⁹, rat anti-DN-cadherin²⁰, mouse anti-β-catenin (Armadillo)²⁷, rabbit anti-βH-spectrin²⁸, mouse anti-Cut²⁹, rabbit anti-BarH1 (ref. 30), rabbit anti-β-galactosidase (Cappel), Cy3/Cy5- (Jackson Research) and FITC- (Promega) conjugated secondary antibodies.

Drosophila genetics

Standard genetic methods were used. Clones of cells misexpressing cadherins were generated by the FLP-out method³¹. The following mutants were used: *Roi*¹⁶, *DE-cadherin*^{shgR69} (ref. 32), *DN-cadherin*^{M19} (ref. 20), *DN-cadherin*^{M12} (ref. 20). The following transgenes were used: *GMR-FLP*³³, *UAS-GFP*, *tub-GAL80 tub-GAL4* (ref. 34), *Ay-GAL4* (ref. 35), *ey-FLP*³⁶, *hs-FLP*³⁷, *arm-LacZ*³⁸, *UAS-DE-cadherin(DEF)*³⁹ and *UAS-DN-cadherin*²⁰. The genotypes of animals that were analysed are as follows (listed by figure): Canton-S and SM6B *Roi*⁺ were used as wild type and *Roi*⁺ mutant strains, respectively (Fig. 2); *y w GMR-FLP/Y; FRT42B DE-cadherin*^{shgR69} *UAS-GFP/ FRT42B tub-GAL80; tub-GAL4/+* (Fig. 3g–i); *y w GMR-FLP/Y; DN-cadherin*^{M19} *FRT42B DE-cadherin*^{shgR69} */ DN-cadherin*^{M12} *FRT42B* (Fig. 3m–r); *y w ey-FLP/+; DN-cadherin*^{M19} *FRT40A / arm-LacZ FRT40A* (Fig. 4); *y w hs-FLP/+; Ay-GAL4 UAS-GFP/UAS-DN-cadherin* (Fig. 5a, c–h). Induction of *hs-FLP* was performed at 34 °C for 30 min at the white-pupa stage. (Fig. 5b) *y w hs-FLP/+; Ay-GAL4 UAS-GFP/UAS-DE-cadherin*. Induction of *hs-FLP* was performed at 34 °C for 30 min at the white-pupa stage.

Received 25 May; accepted 19 August 2004; doi:10.1038/nature02952.

- Etienne-Manneville, S. & Hall, A. Rho GTPases in cell biology. *Nature* **420**, 629–635 (2002).
- Knox, A. L. & Brown, N. H. Rap1 GTPase regulation of adherens junction positioning and cell adhesion. *Science* **295**, 1285–1288 (2002).
- Thompson, D. W. *On Growth and Form* (Cambridge Univ. Press, Cambridge, 1917).
- Plateau, J. A. F. *Statique Experimentale et Theorique des Liquides Soumis aux Seules Forces Moleculaires* (Gauthier-Villars, Paris, 1873).
- Taylor, J. E. The structure of singularities in soap-bubble-like and soap-film-like minimal surfaces. *Ann. Math.* **103**, 489–539 (1976).
- Steinberg, M. S. Reconstruction of tissues by dissociated cells: some morphogenetic tissue movements and the sorting out of embryonic cells may have a common explanation. *Science* **141**, 401–408 (1963).
- Steinberg, M. S. & Takeichi, M. Experimental specification of cell sorting, tissue spreading, and specific spatial patterning by quantitative differences in cadherin expression. *Proc. Natl Acad. Sci. USA* **91**, 206–209 (1994).
- Foty, R. A., Pflieger, C. M., Forgacs, G. & Steinberg, M. S. Surface tensions of embryonic tissues predict their mutual envelopment behavior. *Development* **122**, 1611–1620 (1996).
- Davis, G. S., Phillips, H. M. & Steinberg, M. S. Germ-layer surface tensions and “tissue affinities” in *Rana pipiens* gastrulae: quantitative measurements. *Dev. Biol.* **192**, 630–644 (1997).
- Chichilnisky, E. J. A mathematical model of pattern formation. *J. Theor. Biol.* **123**, 81–101 (1986).
- Ready, D. F., Hanson, T. E. & Benzer, S. Development of the *Drosophila* retina, a neurocrystalline lattice. *Dev. Biol.* **53**, 217–240 (1976).
- Cagan, R. L. & Ready, D. F. The emergence of order in the *Drosophila* pupal retina. *Dev. Biol.* **136**, 346–362 (1989).
- Wolff, T. & Ready, D. F. In *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez-Arias, A.) 1277–1325 (Cold Spring Harbor, New York, 1993).
- Strausfeld, N. J. & Nassel, D. R. in *Handbook of Sensory Physiology* (ed. Autrum, H.) 1–132 (Springer, Berlin, 1981).
- Chanut, F. *et al.* *Rough eye* is a gain-of-function allele of *amos* that disrupts regulation of the proneural gene *atonal* during *Drosophila* retinal differentiation. *Genetics* **160**, 623–635 (2002).
- Renfranz, P. J. & Benzer, S. Monoclonal antibody probes discriminate early and late mutant defects in development of the *Drosophila* retina. *Dev. Biol.* **136**, 411–429 (1989).

- Tepass, U., Tanentzapf, G., Ward, R. & Fehon, R. Epithelial cell polarity and cell junctions in *Drosophila*. *Annu. Rev. Genet.* **35**, 747–784 (2001).
- Jamora, C. & Fuchs, E. Intercellular adhesion, signalling and the cytoskeleton. *Nature Cell Biol.* **4**, E101–E108 (2002).
- Oda, H., Uemura, T., Harada, Y., Iwai, Y. & Takeichi, M. A *Drosophila* homolog of cadherin associated with *armadillo* and essential for embryonic cell-cell adhesion. *Dev. Biol.* **165**, 716–726 (1994).
- Iwai, Y. *et al.* Axon patterning requires DN-cadherin, a novel neuronal adhesion receptor, in the *Drosophila* embryonic CNS. *Neuron* **19**, 77–89 (1997).
- Hynes, R. O. & Zhao, Q. The evolution of cell adhesion. *J. Cell Biol.* **150**, 89F–96F (2000).
- Sanson, B., White, P. & Vincent, J. P. Uncoupling cadherin-based adhesion from *wingless* signalling in *Drosophila*. *Nature* **383**, 627–630 (1996).
- Zhu, B. *et al.* Functional analysis of the structural basis of homophilic cadherin adhesion. *Biophys. J.* **84**, 4033–4042 (2003).
- Duguay, D., Foty, R. A. & Steinberg, M. S. Cadherin-mediated cell adhesion and tissue segregation: qualitative and quantitative determinants. *Dev. Biol.* **253**, 309–323 (2003).
- Bloor, J. W. & Kiehart, D. P. *Drosophila* RhoA regulates the cytoskeleton and cell-cell adhesion in the developing epidermis. *Development* **129**, 3173–3183 (2002).
- Hayashi, T., Kojima, T. & Saigo, K. Specification of primary pigment cell and outer photoreceptor fates by *BarH1* homeobox gene in the developing *Drosophila* eye. *Dev. Biol.* **200**, 131–145 (1998).
- Peifer, M., Pai, L. M. & Casey, M. Phosphorylation of the *Drosophila* adherens junction protein Armadillo: roles for *wingless* signal and *zeste-white* 3 kinase. *Dev. Biol.* **166**, 543–556 (1994).
- Thomas, G. H. *et al.* *Drosophila* β-heavy-spectrin is essential for development and contributes to specific cell fates in the eye. *Development* **125**, 2125–2134 (1998).
- Blochlinger, K., Jan, L. Y. & Jan, Y. N. Postembryonic patterns of expression of *cut*, a locus regulating sensory organ identity in *Drosophila*. *Development* **117**, 441–450 (1993).
- Higashijima, S. *et al.* Dual *Bar* homeo box genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. *Genes Dev.* **6**, 50–60 (1992).
- Basler, K. & Struhl, G. Compartment boundaries and the control of *Drosophila* limb pattern by *hedgehog* protein. *Nature* **368**, 208–214 (1994).
- Godt, D. & Tepass, U. *Drosophila* oocyte localization is mediated by differential cadherin-based adhesion. *Nature* **395**, 387–391 (1998).
- Pignoni, F. & Zipursky, S. L. Induction of *Drosophila* eye development by *decapentaplegic*. *Development* **124**, 271–278 (1997).
- Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
- Ito, K., Awano, W., Suzuki, K., Hiromi, Y. & Yamamoto, D. The *Drosophila* mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. *Development* **124**, 761–771 (1997).
- Newsome, T. P., Asling, B. & Dickson, B. J. Analysis of *Drosophila* photoreceptor axon guidance in eye-specific mosaics. *Development* **127**, 851–860 (2000).
- Golic, K. G. & Lindquist, S. The FLP recombinase of yeast catalyzes site-specific recombination in the *Drosophila* genome. *Cell* **59**, 499–509 (1989).
- Vincent, J. P. & Lawrence, P. A. *Drosophila wingless* sustains engrailed expression only in adjoining cells: evidence from mosaic embryos. *Cell* **77**, 909–915 (1994).
- Oda, H. & Tsukita, S. Nonchordate classic cadherins have a structurally and functionally unique domain that is absent from chordate classic cadherins. *Dev. Biol.* **216**, 406–422 (1999).
- Morgan, T. H. *Experimental Embryology* (Columbia Univ. Press, New York, 1927).

Acknowledgements We thank T. Uemura, H. Oda, U. Tepass, G. Thomas, B. Dickson, P. Garrity, the Bloomington *Drosophila* Stock Center and the Developmental Studies Hybridoma Bank for fly strains and/or antibodies. We also thank K. Saigo for use of facilities. We thank M. Steinberg and R. Matsuda for critical input. We acknowledge G. Beitel, A. Dudley, J. Widom and E. Sontheimer for greatly improving the manuscript. T.H. was supported by a research fellowship from the Japan Society for the Promotion of Science for Young Scientists. This work was also supported by the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.H. (t-hayashi@northwestern.edu) or R.W.C. (r-carthew@northwestern.edu).

Direct integration of *Hox* and segmentation gene inputs during *Drosophila* development

Brian Gebelein¹, Daniel J. McKay² & Richard S. Mann¹

¹Department of Biochemistry and Molecular Biophysics and Center for Neurobiology and Behavior, ²Integrated Program in Cellular, Molecular, and Biophysical Studies, Columbia University, 701 West 168th Street, HHSC 1104, New York, New York 10032, USA

During *Drosophila* embryogenesis, segments, each with an anterior and posterior compartment, are generated by the segmentation genes while the *Hox* genes provide each segment with a unique identity. These two processes have been thought to occur independently. Here we show that abdominal *Hox* proteins work directly with two different segmentation proteins, Sloppy paired and Engrailed, to repress the *Hox* target gene *Distalless* in anterior and posterior compartments, respectively. These results suggest that segmentation proteins can function as *Hox* cofactors and reveal a previously unanticipated use of compartments for gene regulation by *Hox* proteins. Our results suggest that these two classes of proteins may collaborate to directly control gene expression at many downstream target genes.

The segregation of groups of cells into compartments is fundamental to animal development^{1–4}. Originally defined in *Drosophila melanogaster*, compartments are critical for providing cells with their unique positional address^{5,6}. The first compartments to form during *Drosophila* development are the anterior and posterior compartments and the key step to defining them is the activation of the gene *engrailed* (*en*)². Expression of *en*, which encodes a homeodomain transcription factor, results in a posterior compartment fate, and the absence of *en* expression results in an anterior compartment fate^{7,8}. Once activated by gap and pair-rule genes, *en* expression and, consequently, the anterior–posterior compartment boundary later become dependent upon the protein Wingless (*Wg*), which is secreted from adjacent anterior compartment cells^{9,10}. Concurrently with anterior–posterior compartmentalization and segmentation, the expression of the eight *Drosophila Hox* genes is also initially established by the gap and pair-rule genes. The *Hox* genes, however, which also encode homeodomain transcription factors, do not contribute to the formation or number of segments but instead specify their unique identities along the anterior–posterior axis^{11–14}.

This flow of genetic information during *Drosophila* embryogenesis has led to the idea that anterior–posterior compartmentalization and segment identity specification are independent processes^{15–18}. In contrast to this view, we show here that these two pathways are interconnected in previously unrecognized ways. We provide evidence that *Hox* factors directly interact with segmentation proteins such as *En* to control gene expression. Moreover, *Hox* proteins collaborate with two different segmentation proteins in anterior and posterior cell types to regulate the same *Hox* target gene, revealing a previously unknown use of compartments to control gene expression by *Hox* proteins.

Distalless expression straddles the compartment boundary

Distalless (*Dll*) is a *Hox* target gene that is required for leg development in *Drosophila*¹⁹. In each thoracic hemisegment, *wg*, expressed by anterior cells adjacent to the anterior–posterior compartment boundary, activates *Dll* in a group of cells that straddle this boundary^{20,21} (Fig. 1). A cis-regulatory element derived from *Dll*, called DMX, drives accurate *Dll*-like expression in the thorax (Fig. 1 and Supplementary Fig. 1a). The abdominal *Hox* genes *Ultrabithorax* (*Ubx*) and *abdominalA* (*abdA*) directly repress *Dll*

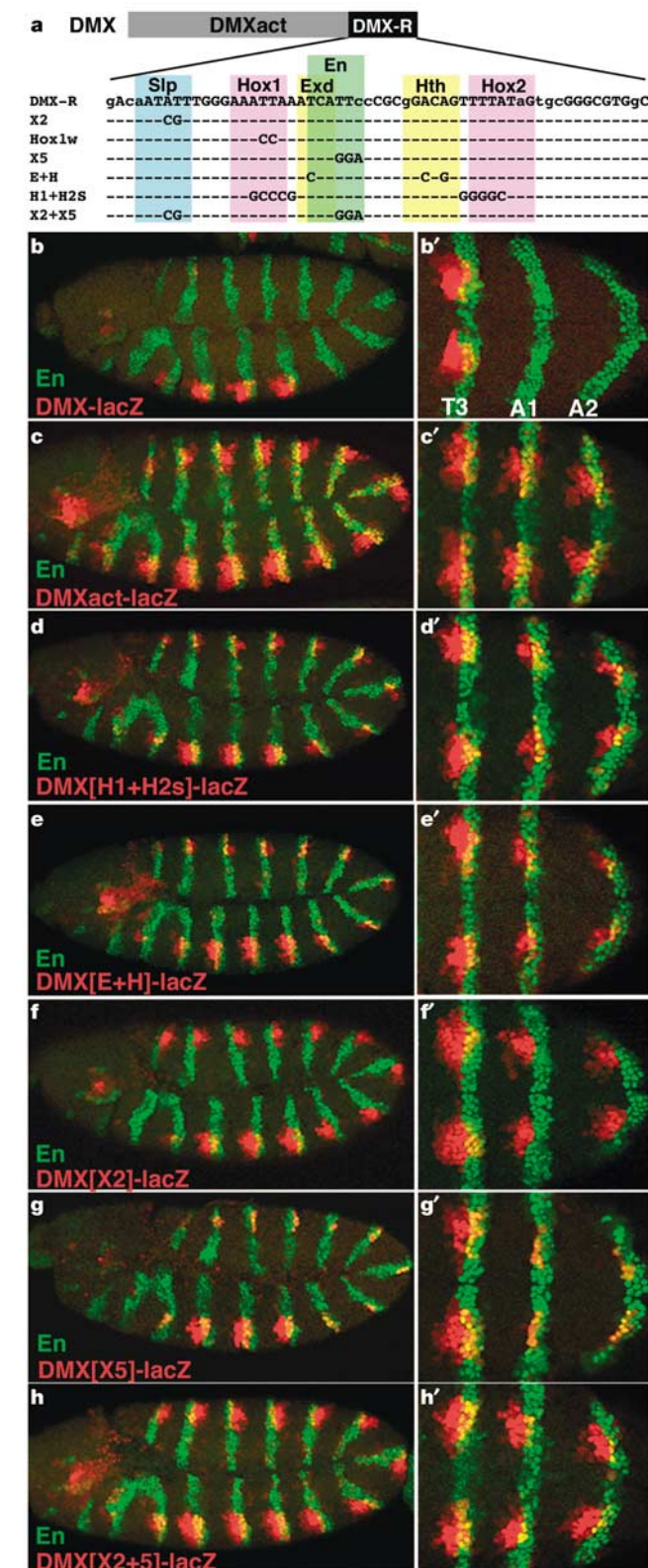
and *DMX-lacZ* in both compartments, thereby blocking leg development in the abdomen^{21–23} (Supplementary Fig. 1b, c). DMX is composed of a large activator element (DMXact) and a 57-base-pair (bp) repressor element referred to here as DMX-R (Fig. 1a–c). Previous work demonstrated that *Ubx* and *abdA* cooperatively bind to DMX-R with two homeodomain cofactors, Extradenticle (*Exd*) and Homothorax (*Hth*)²³. In contrast, the thoracic *Hox* protein Antennapedia (*Antp*) does not repress *Dll* and does not bind DMX-R with high affinity in the presence or absence of *Exd* and *Hth*²³. Thus, repression of *Dll* in the abdomen depends in part on the ability of these cofactors to selectively enhance the binding of the abdominal *Hox* proteins to DMX-R²³.

Compartment-specific *Dll* repression

Exd and *Hth*, as well as their vertebrate counterparts, are used as *Hox* cofactors at many target genes²⁴. Moreover, *Hox/Exd/Hth* complexes are used for both gene activation and repression, raising the question of how the decision to activate or repress is determined. One view posits that these complexes do not directly recruit co-activators or co-repressors, but instead are required for target gene selection²⁵. Accordingly, other DNA sequences present at *Hox/Exd/Hth*-targeted elements would determine whether a target gene is activated or repressed. Consistent with this notion, DMX-R sequences isolated from six *Drosophila* species show extensive conservation outside the previously identified *Hox* (referred to here as *Hox1*) *Exd* and *Hth* binding sites (Supplementary Fig. 1d), suggesting that they also play a role in *Dll* regulation.

To test a role for these conserved sequences, we performed a thorough mutagenesis of DMX-R (Fig. 1; a complete summary of the mutagenesis is shown in Supplementary Fig. 1e). Each mutant DMX-R was cloned into an otherwise wild-type, full-length DMX and tested for activity in a standard reporter gene assay in transgenic embryos. Thoracic expression was normal in all cases. However, to our surprise, many of the DMX-R mutations, such as X5, resulted in abdominal de-repression only in *En*-positive posterior compartment cells, whereas other mutations, such as X2, resulted in abdominal de-repression only in *En*-negative anterior compartment cells (Fig. 1f, g). Single mutations in the *Hox1*, *Exd*, or *Hth* sites also resulted in de-repression predominantly in posterior cells (Supplementary Fig. 1e). In contrast, deletion of the entire DMX-R (*DMXact-lacZ*), or mutations in both the X2 and X5 sites

(DMX[X2 + X5]-lacZ), resulted in de-repression in both compartments (Fig. 1c, h). These results suggest that distinct repression complexes bind to the DMX-R in the anterior and posterior compartments and that segmentation genes play a role in *Dll* repression.



Hox input is mediated by a core Hox/Exd/Hth/Hox complex

One clue to the identity of the proteins in these repression complexes was that the sequence around the Hth site was nearly identical to a Hth/Hox binding site that had been identified previously by a systematic evolution of ligands by exponential enrichment (SELEX) approach using vertebrate Hox and Meis proteins²⁶ (Supplementary Fig. 1f). This similarity suggested the presence of a second, potentially redundant Hox binding site, Hox2. In agreement with this idea, mutations in both the Hox1 and Hox2 binding sites resulted in de-repression in both the anterior and posterior compartments of the abdominal segments (Fig. 1d). Similarly, although individual mutations in the Exd and Hth binding sites lead predominantly to de-repression in the posterior compartment, mutation of both sites resulted in de-repression in both compartments (Fig. 1e). These results suggest that a Hox/Exd/Hth/Hox complex may be used for repression in both compartments. Furthermore, they suggest that although single mutations in these binding sites are sufficient to disrupt the activity of this complex in the posterior compartment, double mutations are required to disrupt its activity in the anterior compartment.

To provide biochemical evidence for a Hox/Exd/Hth/Hox tetramer, we performed DNA binding experiments using DMX-R probes and proteins expressed and purified from *Escherichia coli* (Supplementary Fig. 2a). Previous experiments demonstrated that a Hox/Exd/Hth trimer cooperatively bound to the Hox1, Exd and Hth sites²³ (Supplementary Fig. 2b). We tested the function of the Hox2 site in two ways. First, we measured binding to a probe, DMX-R2, that includes the Exd, Hth and Hox2 sites, but not the Hox1 site (Supplementary Fig. 2c). We found that Exd/Hth/AbdA and Exd/Hth/Ubx trimers cooperatively bound to this probe and that mutations in the Hth, Exd or Hox2 binding sites reduced or eliminated complex formation (Supplementary Fig. 2c and data not shown).

Second, if both the Hox1 and Hox2 sites are functional, the full-length DMX-R may promote the assembly of Hox/Exd/Hth/Hox tetramers. Using a probe containing all four binding sites (DMX-R1+2), we observed the formation of such complexes (Fig. 2a, b). Mutation of any of the four binding sites reduced the amount of tetramer binding whereas mutation of both Hox sites or both the Exd and Hth sites eliminated tetramer binding (Fig. 2b and data not shown). Furthermore, Antp, which does not repress *Dll*, formed tetramers with Exd and Hth that were approximately tenfold weaker than with Ubx or AbdA (Fig. 2c), but bound well to a consensus Hox/Exd/Hth trimer binding site (Fig. 2d). Because mutation of both Hox sites or both the Exd and Hth sites resulted in de-repression in both compartments (Fig. 1), these experiments correlate the binding of a Hox/Exd/Hth/Hox complex on the

Figure 1 Compartment-specific repression mediated by the DMX-R. **a**, Schematic of DMX and sequences of wild type and mutant DMX-Rs. See Supplementary Fig. 1e for the complete list of DMX-R mutants. Upper case letters in the top line indicate base pairs that are 100% conserved in six *Drosophila* species (Supplementary Fig. 1d). **b–h**, Stage-11 embryos containing various DMX-lacZ reporter genes stained for β-gal (red) and En (green). Panels on the left show lateral views of entire embryos; panels on the right show ventral views of the third thoracic segment (T3) to the second abdominal segment (A2). **b**, DMX-lacZ. Reporter expression is limited to the thorax and includes both En+ and En− cells. **c**, DMXact-lacZ. Reporter expression is throughout the embryo and includes both En+ and En− cells. **d**, DMX[H1+H2s]-lacZ. Reporter expression is throughout the embryo and includes both En+ and En− cells. **e**, DMX[E+H]-lacZ. Reporter expression is throughout the embryo and includes both En+ and En− cells. Expression in the abdomen is weaker, however, than DMXact-lacZ. **f**, DMX[X2]-lacZ. In the abdomen, reporter expression is only observed in En− anterior compartment cells. **g**, DMX[X5]-lacZ. In the abdomen, reporter expression is only observed in En+ posterior compartment cells. **h**, DMX[X2+5]-lacZ. Reporter expression is throughout the embryo and includes both En+ and En− cells.

DMX-R with the ability of this element to mediate repression in both compartments.

DMX expression overlaps Slp- or En-expressing cells

Although binding of a Hox/Exd/Hth/Hox tetramer is sufficient to account for the necessary abdominal Hox-input into *Dll* repression, it does not explain the compartment-specific de-repression exhibited by some DMX-R mutations (Fig. 1). The X2 and X5 mutations, for example, result in abdominal de-repression (Fig. 1) but do not prevent the formation of the Hox/Exd/Hth/Hox tetramer (Fig. 2b). Sequence inspection of the DMX-R revealed that the X2 mutation, which resulted in de-repression specifically in the anterior compartment, disrupts two partially overlapping matches to a consensus binding site for Forkhead (Fkh) domain proteins (Supplementary Fig. 1f). With this in mind, we examined the expression pattern of Sloppy paired 1 (Slp1), a Fkh domain factor encoded by one of two partially redundant segmentation genes, *slp1* and *slp2*^{27–29}. The two

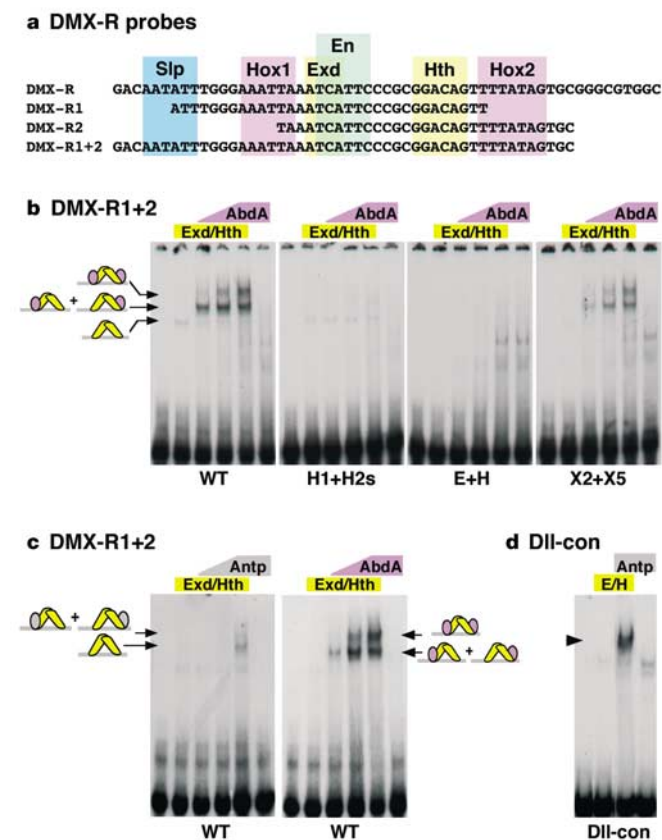


Figure 2 Assembly of a core Hox/Exd/Hth/Hox complex on the DMX-R. **a**, The sequences of full-length DMX-R and three probes used for EMSAs. **b**, **c**, EMSAs using wild type (WT) or mutant versions of the DMX-R1+2 probe. Although these experiments were performed with AbdA, Ubx has similar DNA binding properties (not shown). The complexes formed are indicated on the left using the symbols defined in Fig. 3i. The probes used (indicated below each EMSA) correspond to the mutations listed in Fig. 1a. **b**, On the wild type DMX-R1+2 probe, Exd/Hth, AbdA/Exd/Hth and Exd/Hth/AbdA trimers, and AbdA/Exd/Hth/AbdA tetramers form. Tetramers failed to form on probes with mutations in both the Exd and Hth sites (E+H) or in both Hox sites (H1+H2s). A probe with the X2 and X5 mutations still allowed tetramer binding. The Exdw and single Hox binding site mutations reduced the amount of tetramer formation approximately threefold, but still allowed trimer formation (data not shown). **c**, On the wild type DMX-R1+2 probe, Antp failed to form tetramers and formed trimers at least tenfold more weakly than AbdA. **d**, The same preparation of Antp used in **c** readily forms complexes with Exd/Hth on a consensus probe (Dll-con)²³.

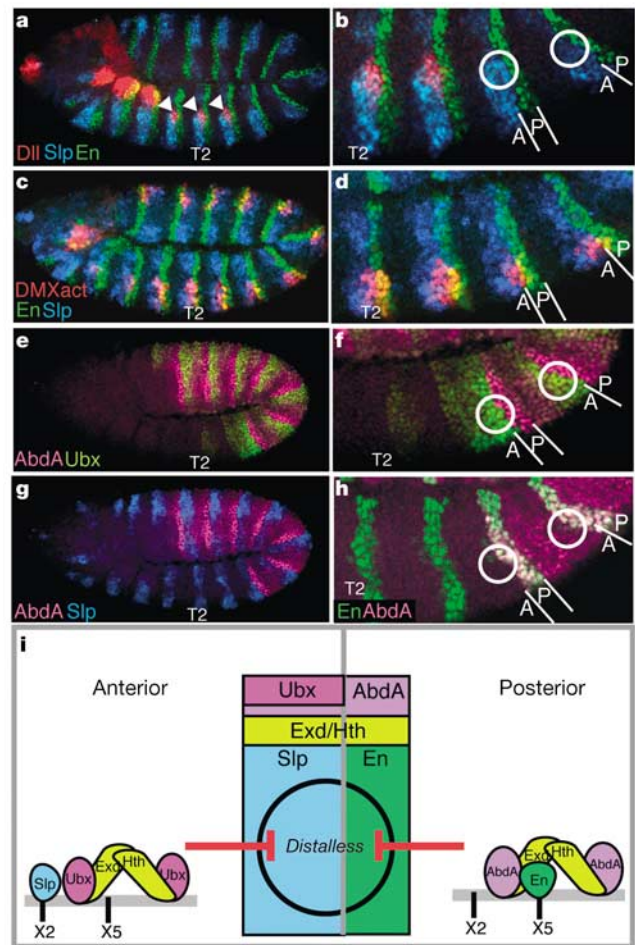


Figure 3 A model for DMX-R-mediated repression of *Dll*. **a–h**, All panels show lateral views of stage-11 embryos. Left-hand panels show entire embryos, right-hand panels show the second thoracic segment (T2) to the second abdominal segment (A2). **a**, **b**, Wild type, stained for *Dll* (red), *Slp* (blue) and *En* (green). The white circles indicate the approximate location of the cells with the potential to express *Dll* in A1 and A2. 'A' and 'P' refer to the anterior and posterior compartments, respectively. In the thorax, *Dll* expression is included within the *Slp* and *En* stripes. In the abdomen, the cells where *Dll* is repressed (white circles) are also included within the *Slp* and *En* stripes. **c**, **d**, *DMXact-lacZ*, stained for β -gal (red), *En* (green) and *Slp* (blue). *DMXact-lacZ* expression in the abdomen marks the cells that have the potential to express *Dll*. These cells are included within the *Slp* and *En* stripes. **e**, **f**, Wild type, stained for *Ubx* (light green) and *AbdA* (pink). In the abdomen, *Ubx* levels are higher in the anterior compartment than in the posterior compartment, whereas *AbdA* levels are higher in the posterior compartment. **g**, Wild type, stained for *AbdA* (pink) and *Slp* (blue). In the abdomen, *AbdA* levels are highest in posterior compartment cells, adjacent to the *Slp*-expressing anterior cells. **h**, Wild type, stained for *AbdA* (purple) and *En* (green). The highest levels of *AbdA* overlap with *En* in the posterior compartment making these nuclei appear white. **i**, Model for DMX-R-mediated repression in the abdomen. In the centre is a summary of the expression patterns of *Ubx*, *AbdA*, *Exd*, *Hth*, *Slp* and *En* in the abdominal segments. In both anterior and posterior compartments we propose that a Hox/Exd/Hth/Hox tetramer binds to the Hox1/Exd/Hth/Hox2 binding sites. On the basis of their expression patterns, *Ubx* and *AbdA* are likely to be the predominant Hox proteins in these complexes in the anterior and posterior compartments, respectively. The model also posits that *Slp* is bound to the X2 site in the anterior compartment and *En* is bound to the X5 site in the posterior compartment. Our data cannot exclude, however, that in the posterior compartment, *En* is bound to the Hox1 site and *AbdA* is bound to the X5 site, or that an *En/Exd/Hth/AbdA* complex is used for repression in the posterior compartment. We favour the scheme shown here because it proposes a similar Hox/Exd/Hth/Hox tetramer in both compartments and it best accommodates the observed cooperative binding between *En* and *AbdA* to the DMX-R (Fig. 4a).

slp genes are expressed in anterior compartment cells adjacent and anterior to En-expressing posterior compartment cells (Fig. 3a–d). In the thorax, cells expressing *Dll* and *DMX-lacZ* co-express either Slp or En at the time *Dll* is initially expressed (Fig. 3a, b and data not shown). In the abdomen, the homologous group of cells, which express *DMXact-lacZ* (a reporter lacking the DMX-R), co-express either Slp in the anterior compartment or En in the posterior compartment (Fig. 3c, d). We also compared the expression patterns of Slp and En with Ubx and AbdA. Ubx levels are highest in anterior, Slp-expressing cells whereas AbdA levels are elevated in posterior, En-expressing cells (Fig. 3e–h). In contrast, both Exd and Hth are present at similar levels in both compartments throughout the abdomen (data not shown).

A model for *Dll* repression in the abdomen

On the basis of these data, we present a model for Hox-mediated repression of *Dll* in both the anterior and posterior compartments of the abdominal segments (Fig. 3i). In the anterior compartment we propose that Slp binds to DMX-R directly with a Ubx/Exd/Hth/Ubx tetramer. In the posterior compartment we suggest that En binds to DMX-R directly with an AbdA/Exd/Hth/AbdA tetramer. One important feature of this model is that Antp/Exd/Hth/Antp complexes fail to form on this DNA, thereby accounting for the lack of repression in the thorax. Furthermore, the model proposes that

Slp and En should, on their own, have only weak affinity for DMX-R sequences because repression does not occur in the thorax, despite the presence of these factors. The Hox/Exd/Hth/Hox complex, perhaps in conjunction with additional factors, is required to recruit or stabilize Slp and En binding to the DMX-R. Both Slp and En are known repressor proteins that directly bind the co-repressor Groucho^{30–33}. Thus, the proposed complexes in both compartments provide a direct link to this co-repressor and, therefore, a mechanism for repression. Below we present DNA binding and genetic experiments that test and support this model.

En and Slp bind DMX-R

To test the idea that En is playing a direct role in *Dll* repression, we examined the ability of En and Hox proteins to bind to DMX-R probes. On its own, En binds to DMX-R very poorly (Fig. 4a–c). Surprisingly, we found that En binds DMX-R with the abdominal Hox proteins Ubx or AbdA in a highly cooperative manner (Fig. 4a, b). The thoracic Hox protein Antp does not bind cooperatively with En to this probe (Fig. 4b). Mutations in the Hox1 or X5 binding sites block AbdA/En binding *in vitro*, consistent with these mutations showing posterior compartment de-repression *in vivo*. In contrast, the X6, X7 and Hth mutations do not affect AbdA/En complex formation (Fig. 4a; Supplementary Figs 1e and 3a).

On the basis of DMX-R's ability to assemble a Hox/Exd/Hth/Hox

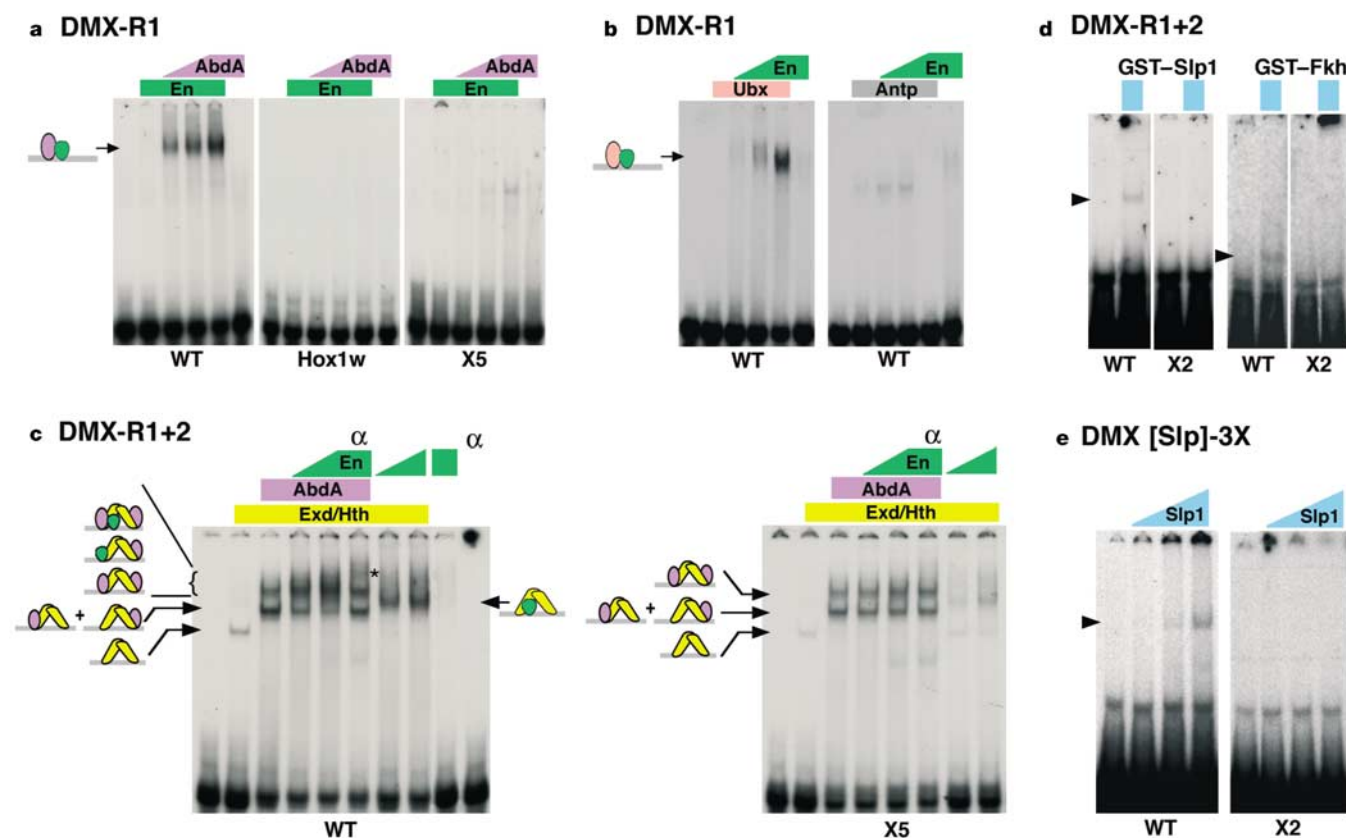


Figure 4 En and Slp bind to DMX-R. **a**, **b**, EMSAs showing that AbdA/En fails to bind probes with mutations in the Hox1 (Hox1w) or X5 sites, which also show posterior compartment de-repression. In contrast, the X6 mutation still allows AbdA/En complex formation (Supplementary Fig. 3). **b**, Ubx/En, like AbdA/En, binds cooperatively to the DMX-R1 probe but Antp does not bind with En in this assay. **c**, Combinations, as indicated, of En, AbdA, Exd and Hth were bound to wild type (WT) or X5 mutant DMX-R1+2 probes. The complexes proposed to form, including En/AbdA/Exd/Hth/AbdA and En/Exd/Hth/AbdA, are indicated. Formation of these higher order complexes is inhibited and partially supershifted (*) by the addition of anti-En antibody (α). On the X5

probe, addition of En does not generate higher order complexes (also see Supplementary Fig. 3b). On the WT probe, an En/Exd/Hth complex is also observed, but this complex is not seen upon addition of AbdA. **d**, **e**, EMSAs showing that Slp1 binds to DMX probes in an X2 dependent manner. **d**, Full-length Slp1 (GST-Slp1) and the forkhead domain of Slp1 (GST-Fkh) bind weakly to the wild type DMX-R1+2 probe (arrowhead) but not to a probe containing the X2 mutation. **e**, GST-Slp1 also binds to a probe containing three copies of the wild type X2 region of DMX-R, but not to a probe containing three copies of the X2 mutant (DMX[Slp]-3x; see Methods).

tetramer (Fig. 2b), we also tested if En could bind together with an AbdA/Exd/Hth/AbdA complex. Addition of En to reactions containing AbdA, Exd and Hth resulted in the formation of a putative En/AbdA/Exd/Hth/AbdA complex (Fig. 4c and Supplementary Fig. 3b). This complex contains En because its formation is inhibited by an anti-En antibody. A weak antibody-induced super-shift is also observed. Moreover, this complex fails to form on the X5 mutant, which causes posterior compartment-specific de-repression (Fig. 4c and Supplementary Fig. 3b). We note that En/Exd/Hth complexes also bind to the DMX-R (Fig. 4c) and that we cannot exclude that an En/Exd/Hth/AbdA complex may be important for *Dll* repression. The model (Fig. 3i) emphasizes a role for an En/AbdA/Exd/Hth/AbdA complex because it better accommodates the cooperative binding observed between En and AbdA on the DMX-R.

Repression in the anterior compartments of the abdominal segments requires the sequence defined by the X2 mutation, which is similar to a Fkh domain consensus binding site (Fig. 1a and Supplementary Fig. 1f). The model predicts that this sequence is bound by Slp (Fig. 3i). Consistent with this view, Slp1 binds weakly to wild type, but not to X2 mutant DMX-R probes (Fig. 4d, e). However, in contrast to En, we do not observe cooperative binding between Slp and Hox or Hox/Exd/Hth/Hox complexes (data not shown), suggesting that additional factors may be required to mediate interactions between Slp and the abdominal Hox factors.

Together, these results suggest that En and Slp play a direct role in *DMX-lacZ* and *Dll* repression. However, these experiments do not unambiguously determine the stoichiometry of binding by these factors. Furthermore, *in vivo*, additional factors may enhance the interaction between these segmentation proteins and Hox complexes, thereby increasing the stability and/or activity of the repression complexes.

Genetic tests of the model

The model for *Dll* repression is supported by previous genetic experiments that examined the effect of *Ubx* and *abdA* mutants on *Dll* expression in the abdomen. *Ubx abdA* double mutants de-repress *Dll* in both compartments of all abdominal segments. In contrast, *Ubx* mutants de-repress *Dll* in the anterior compartment of only the first abdominal segment, which lacks AbdA. *abdA* mutant embryos de-repress *Dll* in the posterior compartments of all abdominal segments, where *Ubx* levels are low^{21,34} (Fig. 3).

We performed several genetic experiments to provide *in vivo* support for the idea that Slp and En work directly with *Ubx* and AbdA to repress *Dll*. The design of these experiments had to take into consideration that the activation of *Dll* in the thorax depends on *wg*, and that *wg* expression depends on both *slp* and *en*²⁹. Consequently, *Dll* expression is mostly absent in *en* or *slp* mutants, making it impossible to characterize the role that these genes play in *Dll* repression from examining *en* or *slp* loss-of-function mutants (data not shown). However, some of the mutant DMX-Rs described here provide the opportunity to test the model in alternative ways.

According to the model, *DMX[X5]-lacZ* is de-repressed in the posterior compartments of the abdominal segments because it fails to assemble the posterior, En-containing complex (Fig. 3i). Repression of *DMX[X5]-lacZ* in the anterior compartments still occurs because it is able to assemble the anterior, Slp-containing complex. According to this model, *DMX[X5]-lacZ* should be fully repressed if Slp is provided in posterior cells. A negative control for this experiment is that ectopic Slp should be unable to repress *DMX[X2]-lacZ* because this reporter gene does not have a functional Slp binding site (Fig. 3i). To mis-express Slp, we used *paired-Gal4* (*prd-Gal4*), which overlaps both the Slp and En stripes in the odd-numbered abdominal segments (data not shown). As predicted, ectopic Slp repressed *DMX[X5]-lacZ* but not *DMX[X2]-lacZ* (Fig. 5b, i), providing strong *in vivo* support for Slp's direct role in *Dll* repression in the anterior compartments.

Conversely, the model posits that *DMX[X2]-lacZ* is de-repressed

in the anterior compartment because it cannot bind Slp, but remains repressed in the posterior compartment because it is able to assemble the En-containing posterior complex (Fig. 3i). Thus, providing En in the anterior compartment should repress *DMX[X2]-lacZ*. A complication with this experiment is that En is a repressor of *Ubx*, which is the predominant abdominal Hox protein in the anterior compartment²¹. We confirmed that *prd-Gal4*-driven expression of En represses *Ubx* and that AbdA levels remain low at the time *Dll* is activated in the thorax (data not shown). Consequently, ectopic En expression is not sufficient to repress *DMX[X2]-lacZ*, consistent with the observation that low levels of abdominal Hox proteins are present (Fig. 5l). Therefore, to promote the assembly of the posterior complex in anterior cells, we co-expressed En with AbdA using *prd-Gal4*. As predicted, this combination of factors repressed *DMX[X2]-lacZ* but not *DMX[X5]-lacZ*, providing strong *in vivo* evidence for En playing an essential role in *Dll* repression in the posterior compartments (Fig. 5f, m).

Several observations provide additional support for the model. First, ectopic expression of AbdA or *Ubx* in the second thoracic segment (T2) represses *DMX[X5]-lacZ* in the anterior compartment, but not in the posterior compartment (Fig. 5d, g). Conversely, expression of AbdA or *Ubx* in T2 represses *DMX[X2]-lacZ* only in posterior compartment cells (Fig. 5k, n). Second, co-expression of Slp with *Ubx* completely represses *DMX[X5]-lacZ* in T2 but does not repress *DMX[X2]-lacZ* in T2 (Fig. 5c, j). Third, in those cases where repression is incomplete (for example, En+AbdA repression of *DMX[X2]-lacZ* in the abdomen), cells that escape repression have low levels of either an abdominal Hox protein or Slp/En (for example, Fig. 5m). Together, these data provide additional evidence that the abdominal Hox proteins work together with Slp and En to repress *Dll*.

Implications of the model

The segregation of cells into anterior and posterior compartments during *Drosophila* embryogenesis is essential for many aspects of fly development^{2,4–6,35}. The results presented here reveal an unanticipated intersection between anterior–posterior compartmentalization by segmentation genes and segment identity specification by Hox genes. Specifically, we suggest that the abdominal Hox proteins collaborate with two different segmentation proteins, Slp and En, to mediate repression of a Hox target gene in the anterior and posterior compartments of the abdomen, respectively. This mechanism of transcriptional repression suggests a previously unknown use of compartments in *Drosophila* development. The mechanism proposed here contrasts with the alternative and simpler hypothesis in which the abdominal Hox proteins would have used the same set of cofactors to repress *Dll* in all abdominal cells, regardless of their compartmental origin.

These results provide further support for the view that Hox/Exd/Hth complexes do not directly bind co-activators or co-repressors but instead indirectly recruit them to regulatory elements. Consistent with previous analyses^{36–39}, we suggest that Hox/Exd/Hth complexes are important for the Hox specificity of target gene selection. Additional factors, such as Slp or En in the case of *Dll* repression, are required to determine whether the target gene will be repressed or activated. In the future, it will be important to dissect in similar detail other Hox-regulated elements, to assess the generality of this mechanism.

These results also broaden the spectrum of cofactors used by Hox proteins to regulate gene expression. Although the analysis of Exd/Hth in *Drosophila* and Pbx/Meis in vertebrates has provided some insights into how Hox specificity is achieved, there are examples of tissues in which these proteins are not available to be Hox cofactors and of Hox targets in which Exd and Hth seem not to play a direct role^{40–44}. We show here that En, a homeodomain segmentation protein, is used as a Hox cofactor to repress *Dll* in the

abdomen. Although the complex defined at the DMX-R includes Exd and Hth, our DNA binding studies demonstrate that Hox and En proteins can bind cooperatively to DNA in the absence of Exd and Hth. These findings suggest that En may function with Ubx and/or AbdA to regulate target genes other than *Dll*, and perhaps independently of Exd and Hth. Consistent with this idea are genetic experiments showing that, in the absence of Exd, En can repress *slp* and this repression requires abdominal *Hox* activity⁴⁵. Although these experiments were unable to distinguish whether the *Hox* input was direct or indirect, our results suggest that En may bind directly

with Ubx and AbdA to repress *slp*, and perhaps other target genes.

Finally, these results raise the question of why a compartment-specific mechanism is used by Hox factors to repress *Dll*. The activation of *Dll* at the compartment boundary by *wg* may be important for accurately positioning the leg primordia within each thoracic hemisegment, but this mode of activation requires that *Dll* is repressed in both compartments in each abdominal segment. The utilization of segmentation proteins such as En and Slp may be the simplest solution to this problem. Compartment-specific mechanisms may also provide additional flexibility in the regulation of

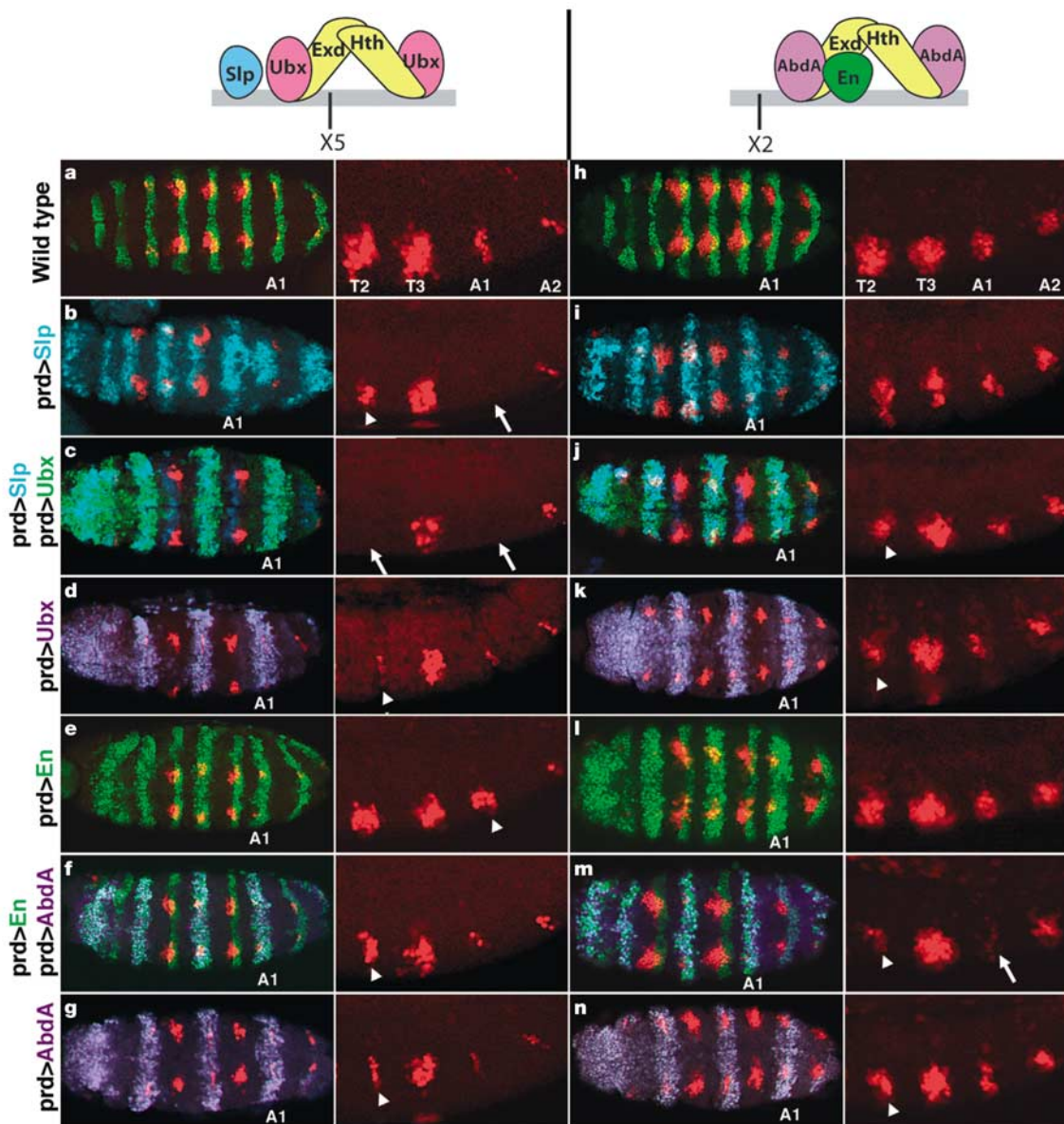


Figure 5 *Dll* is repressed by *Hox* and segmentation gene inputs. Shown on the top are the proposed complexes that can assemble onto the X5 (left) and X2 (right) mutant DMX-Rs. **a–g**, and **h–n**, show *DMX[X5]-lacZ* and *DMX[X2]-lacZ* embryos, respectively, ectopically expressing the indicated proteins via the *prd-Gal4* driver. *prd-Gal4* is expressed in T2 and the odd-numbered abdominal segments. Complete repression of the reporter gene is indicated with arrows, and partial effects are indicated with arrowheads. In all cases, low magnification ventral views are shown on the left and higher magnification lateral views are shown on the right. All embryos were stained for β -gal (red). The ectopically expressed proteins were monitored by antibody staining as indicated. In the embryos with no ectopically expressed proteins (**a** and **h**), the embryos were stained for En (green). In

addition to the effects described in the text, two other results seen in these embryos are noteworthy. When Slp is expressed by itself, we observe partial repression in T2 of *DMX[X5]-lacZ* (arrowhead in **b**), but no effect on *DMX[X2]-lacZ* (**i**). We suggest that when Slp is present at very high levels (as with *prd-Gal4; UAS-Slp*) it can partially repress *DMX[X5]-lacZ* in a Hox-independent manner in the thorax. *DMX[X2]-lacZ* is not repressed, however, because it is unable to bind Slp. Expression of En by itself also resulted in some partial effects. For example, expression of En in A1 caused partial de-repression of *DMX[X5]-lacZ* (arrowhead in **e**) because, we suggest, En represses Slp, which is required for repression of this reporter gene in the anterior compartment.

target genes by Hox proteins by allowing them to turn genes on or off specifically in anterior or posterior cell types. For these reasons, compartment-dependent mechanisms of gene regulation may turn out to be the general rule instead of the exception. □

Methods

Plasmids

The sequences of the DMX-R region from *Drosophila simulans*, *Drosophila teissieri*, *Drosophila erecta*, *Drosophila pseudoobscura* and *Drosophila hydei* were obtained using the polymerase chain reaction (PCR) (details available upon request). On the basis of this analysis the DMX-R was defined from bp 675 to 731 of Dll304²², 18 bp longer than the previously described DME²³, which lacked the Hox2 site. Although shorter than the DMX-R, the DME mediated full repression due to compensating sequences present in adjacent vector sequences (data not shown). However, the truncated element was more sensitive to mutations in the remaining binding sites (for example, Hox1, Exd and Hth), which resulted in de-repression in both compartments. The DMX series of mutations and the DMXact (bps 1 to 680 of Dll304) constructs were generated by the PCR and cloned into the *hs43-nuc-lacZ* vector. All constructs were confirmed by DNA sequencing.

Fly stocks and antibody staining

Expression of *lacZ* (anti- β -gal, Cappel), En (4D9), Ubx (FP3.38), Abd-A⁴⁶, Dll²² and Slp1⁴⁷ were detected by antibody staining and confocal microscopy. The degree of abdominal de-repression of *lacZ* between constructs was normalized to thoracic expression levels. En, Slp1, Ubx and Abd-A misexpression were driven by *prd-Gal4* in the presence of DMX-*lacZ* mutants as indicated.

Protein purification and EMSAs

Hox, Exd and Hth proteins used were purified from BL21 bacteria as His-tagged fusions using Ni-chromatography as described²³. Full-length and the Fkh domain (residues 105–216) of Slp1 were cloned into pGEX5X-1. These proteins were purified from BL21 bacteria using the manufacturer's recommendations (Amersham-Pharmacia). Protein concentrations were measured by the Bradford assay and confirmed by SDS-polyacrylamide gel electrophoresis (SDSPAGE) and Coomassie blue analysis. Electrophoretic mobility shift assays (EMSAs) were performed as previously described²³. In all cases, EMSAs within the same figure panel were performed at the same time and with the same amounts of proteins, so are directly comparable. The amount of protein used in each EMSA was: Fig. 2: Exd/Hth, 50 ng, AbdA and Antp, 15, 45 and 135 ng; Fig. 4b: Ubx and Antp, 40 ng, En, 25, 75 and 225 ng; Fig. 4a: En, 100 ng and AbdA, 15, 45 and 135 ng; Fig. 4c: Exd/Hth, 50 ng, AbdA, 135 ng and En, 75 or 225 ng; Fig. 4d: 500 ng glutathione S-transferase (GST)–Slp1 full-length and 500 ng GST–Fkh; Fig. 4e: 55, 167 or 500 ng GST–Slp1. The probes for Fig. 4e consist of three copies of the wild type (GACAATATTGGGAA) or X2 mutant (GACAATCGTTGGGAA) Slp region of DMX-R (DMX[Slp]-3X). 2 μ l of anti-En antibody (mAb4F11) were used for supershifts.

Received 19 July; accepted 13 August 2004; doi:10.1038/nature02946.

- Garcia-Bellido, A., Ripoll, P. & Morata, G. Developmental compartmentalisation of the wing disk of *Drosophila*. *Nature New Biol.* **245**, 251–253 (1973).
- Morata, G. & Lawrence, P. A. Control of compartment development by the *engrailed* gene in *Drosophila*. *Nature* **255**, 614–617 (1975).
- Fraser, S., Keynes, R. & Lumsden, A. Segmentation in the chick embryo hindbrain is defined by cell lineage restrictions. *Nature* **344**, 431–435 (1990).
- Martínez-Arias, A. & Lawrence, P. A. Parasegments and compartments in the *Drosophila* embryo. *Nature* **313**, 639–642 (1985).
- Irvine, K. D. & Raushkolb, C. Boundaries in development: formation and function. *Annu. Rev. Cell Dev. Biol.* **17**, 189–214 (2001).
- Lawrence, P. A. & Struhl, G. Morphogens, compartments, and pattern: lessons from *Drosophila*?. *Cell* **85**, 951–961 (1996).
- Fjose, A., McGinnis, W. J. & Gehring, W. J. Isolation of a homoeo box-containing gene from the *engrailed* region of *Drosophila* and the spatial distribution of its transcripts. *Nature* **313**, 284–289 (1985).
- Kornberg, T., Siden, I., O'Farrell, P. & Simon, M. The *engrailed* locus of *Drosophila*: in situ localization of transcripts reveals compartment-specific expression. *Cell* **40**, 45–53 (1985).
- Nasiadka, A., Dietrich, B. H. & Krause, H. M. Anterior–posterior patterning in the *Drosophila* embryo. *Adv. Dev. Biol. Biochem.* **12**, 155–204 (2002).
- DiNardo, S., Heemskerck, J., Dougan, S. & O'Farrell, P. H. The making of a maggot: patterning the *Drosophila* embryonic epidermis. *Curr. Opin. Genet. Dev.* **4**, 529–534 (1994).
- Sánchez-Herrero, E., Vernos, I., Marco, R. & Morata, G. Genetic organization of *Drosophila* bithorax complex. *Nature* **313**, 108–113 (1985).
- Carroll, S. B., DiNardo, S., O'Farrell, P. H., White, R. A. & Scott, M. P. Temporal and spatial relationships between segmentation and homeotic gene expression in *Drosophila* embryos: distributions of the fushi tarazu, engrailed, Sex combs reduced, Antennapedia, and Ultrabithorax proteins. *Genes Dev.* **2**, 350–360 (1988).
- McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. *Cell* **68**, 283–302 (1992).
- Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
- Casanova, J., Sanchez-Herrero, E., Busturia, A. & Morata, G. Double and triple mutant combinations of the bithorax complex of *Drosophila*. *EMBO J.* **6**, 3103–3109 (1987).
- Gibson, G. & Gehring, W. J. Head and thoracic transformation caused by ectopic expression of *Antennapedia* during *Drosophila* development. *Development* **102**, 657–675 (1988).
- Mann, R. S. & Hogness, D. S. Functional dissection of Ultrabithorax proteins in *D. melanogaster*. *Cell* **60**, 597–610 (1990).
- Gonzalez-Reyes, A. & Morata, G. The developmental effect of overexpressing a Ubx product in *Drosophila* embryos is dependent on its interactions with other homeotic products. *Cell* **61**, 515–522 (1990).
- Cohen, S. M., Bronner, G., Kuttner, F., Jurgens, G. & Jackle, H. *Distal-less* encodes a homeodomain protein required for limb development in *Drosophila*. *Nature* **338**, 432–434 (1989).
- Cohen, S. M. Specification of limb development in the *Drosophila* embryo by positional cues from segmentation genes. *Nature* **343**, 173–177 (1990).
- Mann, R. *engrailed*-mediated repression of *Ultrabithorax* is necessary for the parasegment 6 identity in *Drosophila*. *Development* **120**, 3205–3212 (1994).
- Vachon, G. *et al.* Homeotic genes of the bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *Distal-less*. *Cell* **71**, 437–450 (1992).
- Gebelein, B., Culi, J., Ryoo, H. D., Zhang, W. & Mann, R. S. Specificity of *Distalless* repression and limb primordia development by abdominal Hox proteins. *Dev. Cell* **3**, 487–498 (2002).
- Mann, R. S. & Carroll, S. B. Molecular mechanisms of selector gene function and evolution. *Curr. Opin. Genet. Dev.* **12**, 592–600 (2002).
- Mann, R. S. & Morata, G. The developmental and molecular biology of genes that subdivide the body of *Drosophila*. *Annu. Rev. Cell Dev. Biol.* **16**, 243–271 (2000).
- Shen, W.-F. *et al.* AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Mol. Cell. Biol.* **17**, 6448–6458 (1997).
- Cadigan, K. M., Grossniklaus, U. & Gehring, W. J. Functional redundancy: the respective roles of the two sloppy paired genes in *Drosophila* segmentation. *Proc. Natl Acad. Sci. USA* **91**, 6324–6328 (1994).
- Grossniklaus, U., Pearson, R. K. & Gehring, W. J. The *Drosophila* sloppy paired locus encodes two proteins involved in segmentation that show homology to mammalian transcription factors. *Genes Dev.* **6**, 1030–1051 (1992).
- Cadigan, K. M., Grossniklaus, U. & Gehring, W. J. Localized expression of sloppy paired protein maintains the polarity of *Drosophila* parasegments. *Genes Dev.* **8**, 899–913 (1994).
- Andrioli, L. P., Vasisht, V., Theodosopoulou, E., Oberstein, A. & Small, S. Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms. *Development* **129**, 4931–4940 (2002).
- Kobayashi, M., Goldstein, R. E., Fujioka, M., Paroush, Z. & Jaynes, J. B. Groucho augments the repression of multiple Even skipped target genes in establishing parasegment boundaries. *Development* **128**, 1805–1815 (2001).
- Tolkunova, E. N., Fujioka, M., Kobayashi, M., Deka, D. & Jaynes, J. B. Two distinct types of repression domain in engrailed: one interacts with the groucho corepressor and is preferentially active on integrated target genes. *Mol. Cell. Biol.* **18**, 2804–2814 (1998).
- Jimenez, G., Paroush, Z. & Ish-Horowitz, D. Groucho acts as a corepressor for a subset of negative regulators, including Hairy and Engrailed. *Genes Dev.* **11**, 3072–3082 (1997).
- Struhl, G. Splitting the bithorax complex of *Drosophila*. *Nature* **308**, 454–457 (1984).
- Dahmann, C. & Basler, K. Compartment boundaries: at the edge of development. *Trends Genet.* **15**, 320–326 (1999).
- Ryoo, H. D. & Mann, R. S. The control of trunk Hox specificity and activity by Extradenticle. *Genes Dev.* **13**, 1704–1716 (1999).
- Pöpperl, H. *et al.* Segmental expression of *Hoxb-1* is controlled by a highly conserved autoregulatory loop dependent upon *exd/Pbx*. *Cell* **81**, 1031–1042 (1995).
- Ferretti, E. *et al.* Segmental expression of Hoxb2 in r4 requires two separate sites that integrate cooperative interactions between Prep1, Pbx and Hox proteins. *Development* **127**, 155–166 (2000).
- Grieder, N., Marty, T., Ryoo, H.-D., Mann, R. S. & Affolter, M. Synergistic activation of a *Drosophila* enhancer by HOM/EXD and DPP signaling. *EMBO J.* **16**, 7402–7410 (1997).
- Rieckhof, G., Casares, F., Ryoo, H. D., Abu-Shaar, M. & Mann, R. S. Nuclear translocation of Extradenticle requires *homothorax*, which encodes an Extradenticle-related homeodomain protein. *Cell* **91**, 171–183 (1997).
- Galant, R., Walsh, C. M. & Carroll, S. B. Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* **129**, 3115–3126 (2002).
- Mann, R. & Abu-Shaar, M. Nuclear import of the homeodomain protein Extradenticle in response to Decapentaplegic and Wingless signalling. *Nature* **383**, 630–633 (1996).
- Aspland, S. E. & White, R. A. Nucleocytoplasmic localisation of extradenticle protein is spatially regulated throughout development in *Drosophila*. *Development* **124**, 741–747 (1997).
- Lohmann, I., McGinnis, N., Bodmer, M. & McGinnis, W. The *Drosophila* Hox gene deformed sculpts head morphology via direct regulation of the apoptosis activator reaper. *Cell* **110**, 457–466 (2002).
- Alexandre, C. & Vincent, J. P. Requirements for transcriptional repression and activation by Engrailed in *Drosophila* embryos. *Development* **130**, 729–739 (2003).
- Karch, F., Bender, W. & Weiffenbach, B. abdA expression in *Drosophila* embryos. *Genes Dev.* **4**, 1573–1587 (1990).
- Kosman, D., Small, S. & Reinitz, J. Rapid preparation of a panel of polyclonal antibodies to *Drosophila* segmentation proteins. *Dev. Genes Evol.* **208**, 290–294 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank K. Cadigan, J. Jaynes, H. Krause, J. Reinitz, C. Schwartz, S. Small, G. Struhl and the Developmental Studies Hybridoma Bank, University of Iowa, for reagents; W. Zhang for technical help; and R. Axel, J. Culi, J. Dasen, O. Hobert, T. Jessell, L. Johnston and G. Struhl for critically reading and commenting on the manuscript. This work was supported by an NIH grant to R.S.M.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.S.M. (rsm10@columbia.edu).

An early extrasolar planetary system revealed by planetesimal belts in β Pictoris

Yoshiko Katata Okamoto^{1,2*}, Hirokazu Katata², Mitsuhiro Honda³, Takuya Yamashita⁴, Takashi Onaka³, Jun-ichi Watanabe⁵, Takashi Miyata⁶, Shigeyuki Sako⁶, Takuya Fujiyoshi⁴ & Itsuki Sakon³

¹Center for Natural Sciences, College of Liberal Arts and Sciences, Kitasato University, 1-15-1 Kitazato, Sagami-hara, Kanagawa 228-8555, Japan

²Department of Infrared Astrophysics, Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Sagami-hara, Kanagawa 229-8510, Japan

³Department of Astronomy, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

⁴Subaru Telescope, National Astronomical Observatory of Japan, 650 North Aohoku Place, Hilo, Hawaii 96720, USA

⁵National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

⁶Institute of Astronomy, School of Science, University of Tokyo, 2-21-1 Osawa, Mitaka, Tokyo 181-0015, Japan

* Present address: Institute of Astrophysics and Planetary Sciences, Ibaraki University, Bunkyo 2-1-1, Mito, Ibaraki 310-8512, Japan

β Pictoris (β Pic) is a main-sequence star with an edge-on dust disk^{1–3} that might represent a state of the early Solar System. The dust does not seem to be a remnant from the original protoplanetary disk, but rather is thought to have been generated from large bodies like planetesimals and/or comets^{4,5}. The history and composition of the parent bodies can therefore be revealed by determining the spatial distribution, grain size, composition and crystallinity of the dust through high-resolution mid-infrared observations. Here we report that the sub-micrometre amorphous silicate grains around β Pic have peaks in their distribution around 6, 16 and 30 μ m (1 AU is the Sun–Earth distance), whereas the crystalline and micrometre-sized amorphous silicate grains are concentrated in the disk centre. As sub-micrometre grains are blown quickly out from the system by radiation pressure from the central star, the peaks indicate the locations of ongoing dust replenishment, which originates from ring-like distributions of planetesimals or ‘planetesimal belts’.

β Pic is a nearby (19.28 pc = 63 light yr)⁶, young (20 Myr)⁷, A5V Vega-like star. A high-resolution 17.9- μ m image of its inner disk (radius $r < 100$ AU) has revealed the existence of dust rings at 14, 28, 52 and 82 AU (ref. 8). The locations of the rings correspond to the region inside the Edgeworth–Kuiper belt (50 AU) of our Solar System, where the planets are formed. The rings indicate a planetary system containing multiple planets, though the nature of the rings remains an open question. During planetary system formation, the materials in protoplanetary disks experience substantial processing. Whereas silicate grains are observed as amorphous in interstellar space, partially crystallized silicate grains are sometimes found in protoplanetary disks^{9,10} and comets¹¹. Dust emission of β Pic also shows the 9.7- μ m amorphous silicate feature and the 11.2- μ m crystalline olivine ((Mg,Fe)₂SiO₄) feature, like some comets¹². Coexistence of crystalline silicates with icy cold materials in comets is paradoxical, because the crystallization of amorphous silicates by annealing requires high temperatures (>1,000 K)¹³. Investigations of the locations of amorphous and crystalline silicates in β Pic provide key information to help the understanding of the crystallization and evolution of dust grains in an early stage of the planetary disk.

We made 10- μ m-band spectroscopic observations ($\lambda/\Delta\lambda \approx 250$, where λ indicates wavelength) of β Pic with the Cooled Mid-Infrared Camera and Spectrometer (COMICS)^{14–16} on the 8.2-m

Subaru Telescope with high spatial resolution. The obtained spectra are shown in Fig. 1. The spectra for the central part (Fig. 1j–l) are clearly different from the other spectra. They have emission peaks at 10.05, 10.30, 10.40, 11.05 and 11.90 μ m. Three of them match with the Mg-pure crystalline olivine (forsterite, Mg₂SiO₄) peaks (10.06, 10.42 and 11.89 μ m)¹⁷. The 11.24- μ m peak of forsterite in laboratory data¹⁷ can be shifted to 11.1 μ m by the shape effect. The spectra between 8.0 and 13.2 μ m were fitted by a least-squares method with model spectra composed of 0.1- and 2- μ m-radii glassy olivine¹⁸, crystalline forsterite¹⁷, and power-law continuum emission. The 0.1- μ m and 2.0- μ m grains represent small (grain radius $s \leq 1 \mu$ m) and large ($s \approx$ a few μ m) grains, respectively. The power-law emission takes account of the residual emission, and the index is given as a free parameter. The derived spatial distribution of each dust emission (Fig. 2) indicates that the 2- μ m glassy olivine and the crystalline forsterite are concentrated in the disk centre. In contrast, the 0.1- μ m glassy olivine does not exhibit a central concentration, but has distribution peaks at 6.4 AU in both sides with weak humps around 16 and 29 AU in the northeast side and around 32 AU in the southwest side. Even when the residual continuum is fitted with a grey-body dust emission (emissivity $\epsilon \approx \text{const.}$) or dust emission of $\epsilon \propto 1/\lambda$, the peaks of the 0.1- μ m glassy olivine at 6.4 and 16 AU are unambiguously present. Only in the case of the $\epsilon \propto 1/\lambda$ emissivity are the humps around 30 AU not seen clearly.

In the β Pic system, grains smaller than 1.5 μ m should be blown out, because the ratio of radiation pressure to gravity (β) exceeds unity according to calculation¹⁹ and revised luminosity (8.7 times solar luminosity)⁶. The fact that a large amount of small grains exist around 6.4 AU indicates that small grains are being replenished at this radius at present. The weak distribution peaks at 16 and 30 AU suggest that these two radii are also places of ongoing replenishment of small grains. They correspond to the dust rings seen in 17.9 μ m at 14 and 28 AU (ref. 8).

If the grains are replenished by planetesimal collisions, it is likely that not only small grains but also large grains are simultaneously replenished at the same radii. However, grains of a few micrometres in size fall towards the central star by Poynting–Robertson drag, and are concentrated in the disk centre (grain density $n(r) \propto r^{-1}$) in a steady state²⁰. Convolution of the point spread function (PSF) with the disk emission distribution derived by assuming grain surface density scales as $\Sigma(r) \propto (1/r)r \approx r^0$ and the temperature profile in Fig. 2 accounts for the observed PSF-like distribution of the 2- μ m grains fairly well.

Infalling large amorphous grains will be heated to a temperature high enough to be crystallized near the star by stellar radiation. Crystalline forsterite can be produced by such thermal annealing. Grains of 2- μ m size reach a classical annealing temperature of 1,000 K (ref. 13) at 0.3 AU from the star. A recent paper reported the activation energy required to crystallize Mg₂SiO₄ smokes, suggesting that the crystallization can occur in 3 yr at 800 K (ref. 21); this annealing temperature can be obtained at 0.6 AU. The unresolved profile of the crystalline forsterite (a_3 , defined in Fig. 2 legend) indicates that crystalline forsterite is concentrated in a very compact central region. The upper limit of the crystallinity of the silicate grains replenished at 6.4 AU is estimated as 10% (2 σ level), which seems rather low to explain the observed crystalline mass discussed below. These two facts support annealing by the central star. The spectrum taken by the Infrared Space Observatory (ISO) does not show any 20–40- μ m crystalline silicate features²², suggesting that the temperature of crystalline grains (if they exist) is no lower than 500 K, otherwise the 20–40- μ m features would become more than one-third of the 10- μ m features (in Jy). Because the mass absorption coefficient of crystalline forsterite in the 3- μ m region is less than that at 11.24 μ m by more than two orders of magnitude, emission at 3 μ m from forsterite grains is estimated to be less than 0.08 Jy. It is consistent with the ISO spectrum.

What is the replenishing source for the grains around 6.4, 16 and

30 AU? The most likely process is collisions among planetesimals in planetesimal belts similar to the asteroid belt in our Solar System (Fig. 3). The masses of 0.1- μm dust in the belts at 6.4, 16 and 30 AU are all similar to each other (1×10^{16} kg for each of the 6.4, 16 and 30 AU emissions in the northeast side, and 7×10^{16} kg for the whole southwest side). It suggests that three belts supply grains almost equally. The asymmetry of the 6.4-AU small grain enhancement indicates that the 6.4-AU belt is slightly inclined compared to the outer disk (Fig. 3).

Alternatively, evaporation from cometary bodies infalling towards the star could also be a grain replenishment source. It is,

however, difficult to explain by this process the depletion of small grains in the disk centre, the grain replenishment at 16 and 30 AU, and the 0.1- μm grain enhancement at the southwest side at 6.4 AU. Also, grains with large-eccentricity orbits typical of comets are more easily blown out than grains with circular orbits, even if the β value is small. Replenishment in the planetesimal belts is thus the most likely dust supply mechanism in β Pic.

The discovered planetesimal belts may be formed by the gravitational perturbation of planetesimals in unstable orbits due to resonance with possible large planets. In the Solar System, the mean motion resonance plays an important role in the relation

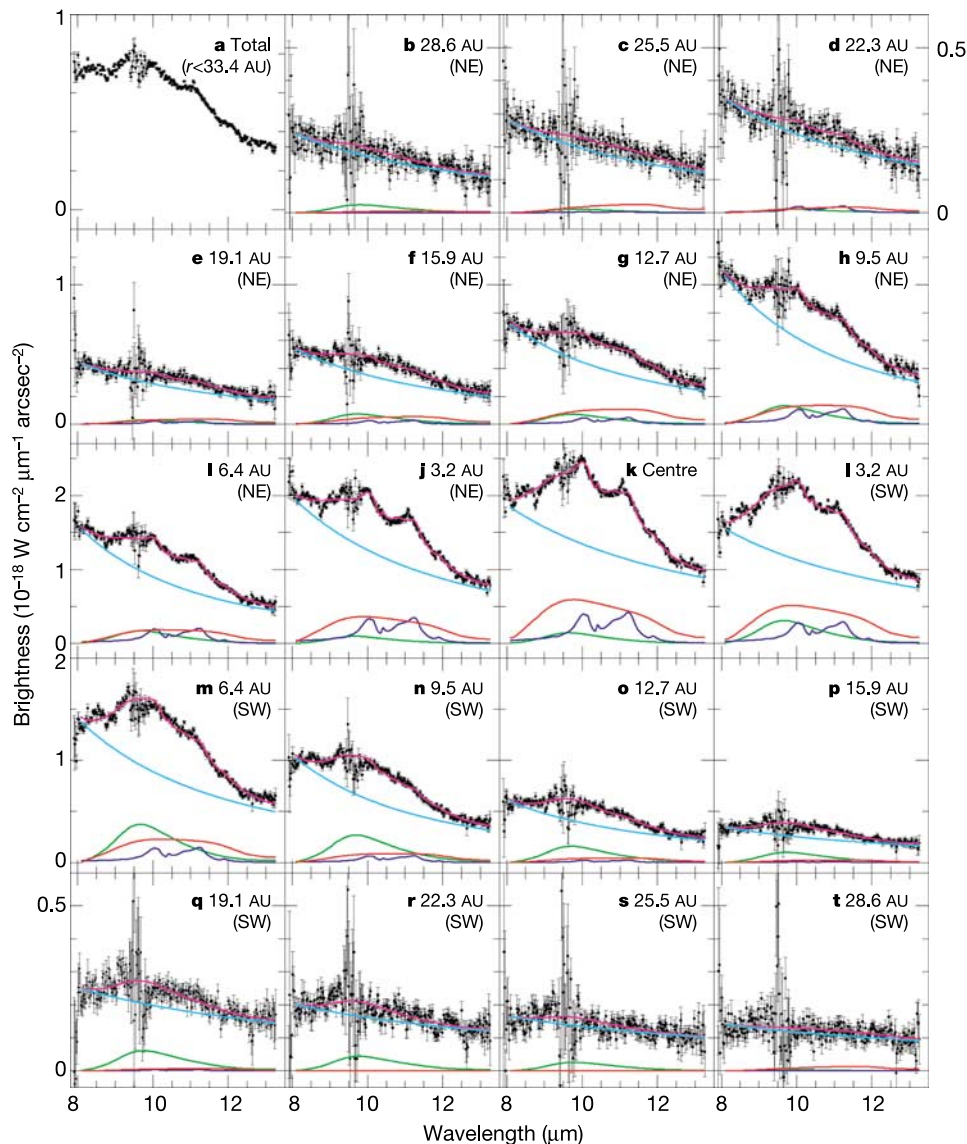


Figure 1 Spectra of β Pic after subtraction of the photospheric emission. **a**, The spectrum integrated along the slit for the $r < 33.4$ AU region. It shows a peak at 11.1 μm and ~ 9.7 μm and a hump around 12 μm seen in past observations¹². **b–t**, The spectrum for each spatial pixel within 26 AU. The position within the disk is indicated in each panel. Northeast and southwest directions are indicated as NE and SW, respectively. The black points with error bars (s.d.) indicate the observed spectra. The atmospheric ozone absorption significantly degrades the signal to noise (S/N) ratio between 9.3 and 9.9 μm . The magenta lines are fitted spectra, and the other coloured lines indicate the contribution from each component: 0.1- μm radius amorphous silicate (green), 2.0- μm radius amorphous silicate (red), crystalline forsterite (blue), and power-law continuum component (cyan). The fitting formula is given in the legend of Fig. 2. The 11.05- μm peak

in the spectra of the disk centre is an unidentified dust feature or line emission overlying the 11.1- μm peak. The observations were long-slit spectroscopy, made on 11 and 12 December 2003 (UT). We used the 0.33"-width slit with a position angle of 30°, almost along the outer disk. The pixel scale was 0.165". The stellar component is subtracted assuming a 8,200 K blackbody and using the PSF from the standard stars, whose full-width at half-maximum is 0.8". Among all the data acquired in more than 3 h of integration time, we have selected 2,983-s data taken under good seeing conditions to obtain a spatial resolution as high as possible. For this selection, the spatial profile of each data file was fitted with a gaussian, and data with the standard deviation of the gaussian no larger than 0.4" were used.

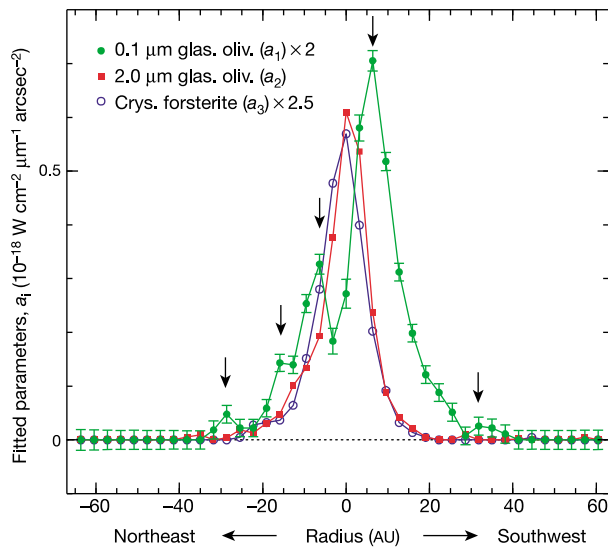


Figure 2 Distribution of each dust emission. The spatial profile of each fit parameter, a_i , which is the brightness of each dust component at $10\ \mu\text{m}$ scaled by each normalized mass absorption coefficient, is shown. The fitting formula is given by $F(\lambda) = a_0(\lambda/10\ \mu\text{m})^n + \sum_{i=1}^3 a_i \kappa_i(\lambda) B(\lambda, T_i) / B(10\ \mu\text{m}, T_i)$, where $F(\lambda)$ is the observed flux at wavelength λ (μm) in $10^{-18}\ \text{W cm}^{-2}\ \mu\text{m}^{-1}\ \text{arcsec}^{-2}$, a_i and κ_i are the normalization factors and the mass absorption coefficients normalized as in our recent paper¹⁰, respectively. The normalizing constants in $\text{cm}^2\ \text{g}^{-1}$ for the 0.1- and 2.0- μm glassy olivine and crystalline forsterite are 2,372.3, 1,664.6 and 2,476.6, respectively. The suffix i indicates three dust species: 0.1- μm radius glassy olivine¹⁸ (1; green), 2- μm radius glassy olivine¹⁸ (2; red), and crystalline forsterite¹⁷ (3; blue). Typical errors for a_2 and a_3 are 0.008 and 0.005 ($10^{-18}\ \text{W cm}^{-2}\ \mu\text{m}^{-1}\ \text{arcsec}^{-2}$), respectively. The errors are in s.e. The continuum emission in the spectra is expressed in the power law of index n with the normalization factor a_0 . The temperatures of amorphous grains^{24,8} are assumed to be given by $T_i(\text{K}) = 468 (L/L_\odot)^{0.2} r^{-0.4}$, where L is the luminosity of the central star (8.7 times the solar luminosity)⁹, s_i is the grain radius in μm , and r is the disk radius in AU at the inner side of each pixel. Their temperatures are assumed to be 700 K at the central pixel. The temperature of crystalline forsterite is assumed to be 1,200 K, since the spatial profile is similar to the PSF (only the width of the profile is about 0.8 times of that of the PSF). Black arrows indicate the distribution peaks of 0.1- μm amorphous silicate grains and the detection of enhancements around 30 AU is marginal. Even if we use the derived distribution of a_3 as the PSF instead of the standard star's profile to subtract the stellar component, the resultant distributions of the components are not changed.

between the planets and the debris belts, such as the asteroid belt and the Edgeworth–Kuiper belt (EKB). The outer and inner edges of the asteroid belt are in 2:1 and 4:1 resonance relations with Jupiter, respectively. If the same relations hold for the 6.4-AU belt around β Pic, the semi-major axis of the planet in resonance and the width of the 6.4-AU belts are estimated as 12 AU and 3 AU, respectively. The latter value is consistent with the belt width unresolved in our observations. The planet at 12 AU is also in a 3:2 resonance relation with the 16-AU belt—like the relation between Neptune and plutinos in the EKB objects.

By assuming that replenished small grains are blown out within the free-fall time (a few to several tens of years), their replenishment rate from each belt is estimated as about 10^{15} – $10^{16}\ \text{kg yr}^{-1}$. This is 10^5 – 10^6 times larger than the replenishment rate of the zodiacal dust²³, and there must be a much larger number of planetesimals than asteroids in the present Solar System. The β Pic system is likely to be at a violent planetesimal collision stage in an early solar system.

To estimate the mass of the 2- μm amorphous and crystalline grains, we assume that the grains are supplied at 6.4 AU as amorphous grains, falling by Poynting–Robertson drag in a steady condition, all being crystallized within the crystallization radius

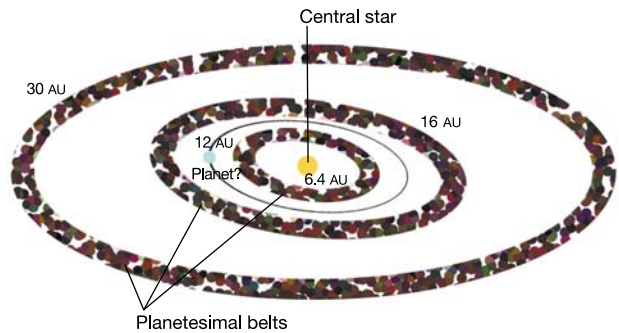


Figure 3 Conceptual view of the discovered planetesimal belts. The locations of the belts are estimated from the dust distribution in Fig. 2. The coloured small dots indicate the planetesimals making up the belts. Planetesimals in the three belts are replenishing amorphous silicate grains by their collisions. The 6.4-AU belt seems to be inclined compared to the outer disk, as the small grain enhancement around 6.4 AU is larger for the southwest side than for the northeast side (Fig. 2). This asymmetry can be explained by a combination of the inclined position angle of the 6.4-AU belt and the present slit position slightly shifted towards the southeast. It is supported by the fact that the inner disk ($r < 20\ \text{AU}$) has been found to be inclined against the outer disk from the mid-infrared images²⁵. The inclination of the 16-AU belt against the outer disk is derived from the 14-AU dust ring in the 17.9- μm image⁸.

r_c , and evaporated at 0.11 AU, where grains reach a temperature of 1,500 K. If the crystallization temperature is assumed to be 800 K ($r_c = 0.6\ \text{AU}$)²¹, the estimated dust masses of the (2- μm) amorphous and crystalline grains are respectively $3 \times 10^{17}\ \text{kg}$ and $7 \times 10^{15}\ \text{kg}$. Taking account of the uncertainties in the assumed temperature profile (Fig. 2) as well as in the density distribution, this suggests that it is reasonable to surmise that the crystalline grains are formed from the amorphous grains near the central star by the stellar radiation. Detailed self-consistent models are required for more precise discussions. □

Received 9 April; accepted 17 August 2004; doi:10.1038/nature02948.

1. Gillett, F. C. in *Light on Dark Matter* (ed. Israel, F. P.) 61–69 (Reidel, Dordrecht, 1986).
2. Smith, B. A. & Terrell, R. J. A circumstellar disk around Beta Pictoris. *Science* **226**, 1421–1424 (1984).
3. Pantin, E., Lagage, P. O. & Artymowicz, P. Mid-infrared images and models of the beta Pictoris dust disk. *Astron. Astrophys.* **327**, 1123–1136 (1997).
4. Backman, D. E. & Paresce, F. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 1253–1304 (Univ. Arizona Press, Tucson, 1993).
5. Lecavelier des Etangs, A., Vidal-Madjar, A. & Ferlet, R. Dust distribution in disks supplied by small bodies: is the β Pictoris disk a gigantic multi-cometary tail? *Astron. Astrophys.* **307**, 542–550 (1996).
6. Crifo, F., Vidal-Madjar, A., Lallemand, R., Ferlet, R. & Gerbaldi, M. β Pictoris revisited by Hipparcos. *Star properties. Astron. Astrophys.* **320**, L29–L32 (1997).
7. Barrado y Navascués, D., Stauffer, J. R., Song, I. & Caillault, J.-P. The age of beta Pictoris. *Astrophys. J.* **520**, L123–L126 (1999).
8. Wahhaj, Z. *et al.* The inner rings of β Pictoris. *Astrophys. J.* **584**, L27–L31 (2003).
9. Malfait, K. *et al.* The spectrum of the young star HD100546 observed with the Infrared Space Observatory. *Astron. Astrophys.* **332**, L25–L28 (1998).
10. Honda, M. *et al.* Detection of crystalline silicates around T Tauri star Hen3–600A. *Astrophys. J.* **585**, L59–L63 (2003).
11. Campins, H. & Ryan, E. V. The identification of crystalline olivine in cometary silicates. *Astrophys. J.* **341**, 1059–1066 (1989).
12. Knacke, R. F. *et al.* The silicates in the disk of beta Pictoris. *Astrophys. J.* **418**, 440–450 (1993).
13. Hallenbeck, S. L., Nuth, J. A. III & Nelson, R. N. Evolving optical properties of annealing silicate grains: from amorphous condensate to crystalline mineral. *Astrophys. J.* **535**, 247–255 (2000).
14. Katata, H. *et al.* COMICS: the cooled mid-infrared camera and spectrometer for the Subaru telescope. *Proc. SPIE* **4008**, 1144–1152 (2000).
15. Okamoto, Y. K. *et al.* Improved performances and capabilities of the Cooled Mid-Infrared Camera and Spectrometer (COMICS) for the Subaru Telescope. *Proc. SPIE* **4841**, 169–180 (2003).
16. Sako, S. *et al.* Improvements in operating the Raytheon 320x240 pixel Si:As impurity band conduction mid-infrared array. *Publ. Astron. Soc. Pacif.* **115**, 1407–1418 (2003).
17. Koike, C. *et al.* Compositional dependence of infrared absorption spectra of crystalline silicate II. Natural and synthetic olivines. *Astron. Astrophys.* **399**, 1101–1107 (2003).
18. Dorschner, J., Begemann, B., Henning, Th., Jäger, C. & Mutschke, H. Steps toward interstellar silicate mineralogy. II. Study of Mg-Fe-silicate glasses of variable composition. *Astron. Astrophys.* **300**, 503–520 (1995).

19. Artymowicz, P. Radiation pressure forces on particles in the Beta Pictoris system. *Astrophys. J.* **335**, L79–L82 (1988).
20. Mukai, T. & Giese, R. H. Modification of the spatial distribution of interplanetary dust grains by Lorentz forces. *Astron. Astrophys.* **131**, 355–363 (1984).
21. Fabian, D., Jäger, C., Henning, Th., Dorschner, J. & Mutschke, H. Steps toward interstellar silicate mineralogy V. Thermal evolution of amorphous magnesium silicates and silica. *Astron. Astrophys.* **364**, 282–292 (2000).
22. Pantin, E., Waelkens, C. & Malfait, K. *The Universe as Seen by ISO* (eds Cox, P. & Kessler, M. F.) 385–388 (ESA SP 427, ESA Publications Division, Noordwijk, 1999).
23. Fixsen, D. J. & Dwek, E. The zodiacal emission spectrum as determined by COBE and its implications. *Astrophys. J.* **578**, 1009–1014 (2002).
24. Backman, D. E., Witteborn, F. C. & Gillett, F. C. Infrared observations and thermal models of the Beta Pictoris disk. *Astrophys. J.* **385**, 670–679 (1992).
25. Weinberger, A. J., Becklin, E. E. & Zuckerman, B. The first spatially resolved mid-infrared spectroscopy of β Pictoris. *Astrophys. J.* **584**, L33–L37 (2003).

Acknowledgements This Letter is based on data collected at the Subaru Telescope, which is operated by the National Astronomical Observatory of Japan. We thank C. Koike, I. Yamamura, F. Usui, S. Hasegawa, T. Ootsubo, H. Chihara, T. Nakamoto, H. Tanaka and T. Takeuchi for comments and discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.K.O. (okamoto@cc.nao.ac.jp).

Magnesium sulphate salts and the history of water on Mars

David T. Vaniman¹, David L. Bish², Steve J. Chipera¹, Claire I. Fialipis¹, J. William Carey¹ & William C. Feldman³

¹Los Alamos National Laboratory (LANL), MS D462, Los Alamos, New Mexico 87545, USA

²Indiana University, 1001 E 10th St, Bloomington, Indiana 47405, USA

³Los Alamos National Laboratory (LANL), MS D466, Los Alamos, New Mexico 87545, USA

Recent reports of ~30 wt% of sulphate within saline sediments on Mars^{1,2}—probably occurring in hydrated form³—suggest a role for sulphates in accounting for equatorial H₂O observed in a global survey by the Odyssey spacecraft⁴. Among salt hydrates likely to be present³, those of the MgSO₄·nH₂O series have many hydration states. Here we report the exposure of several of these phases to varied temperature, pressure and humidity to constrain their possible H₂O contents under martian surface conditions. We found that crystalline structure and H₂O content are dependent on temperature–pressure history, that an amorphous hydrated phase with slow dehydration kinetics forms at <1% relative humidity, and that equilibrium calculations may not reflect the true H₂O-bearing potential of martian soils. Mg sulphate salts can retain sufficient H₂O to explain a portion of the Odyssey observations⁵. Because phases in the MgSO₄·nH₂O system are sensitive to temperature and humidity, they can reveal much about the history of water on Mars. However, their ease of transformation implies that salt hydrates collected on Mars will not be returned to Earth unmodified, and that accurate *in situ* analysis is imperative.

Direct evidence for sulphate salts on Mars dates back to recognition of an Mg–S correlation in X-ray fluorescence data collected by Viking in 1976 at two widely separated landing sites⁶. The Viking data led to the concept of Mg sulphate salt as a widespread cementing agent at ~10 wt% (anhydrous basis) in martian soils⁷, particularly as a shallow ‘duricrust’ within the upper few centimetres of soil. The concept of widespread Mg sulphate salt distribution in soils was reinforced by detection of similar Mg–S

correlations at the Pathfinder landing site in 1997 (refs 8, 9), and by recent determination of comparable Mg–S compositions in soils and within a possible coating of cemented dust on the rock Mazatzal at Gusev crater¹⁰. Global thermal inertia data¹¹ and thermal emission spectroscopy¹² suggest that duricrust may cover vast tracts of land at the martian equator. Widespread distribution of sulphate salt as a cementing agent in soil may be attributed to a relatively young pedogenic process, possibly acidic weathering, that does not require surface water or ground water^{13,14}. A pedogenic origin remains plausible for this particular mode of occurrence, but results from the Mars Exploration Rover Opportunity at Meridiani Planum^{1,2} provide compelling evidence that sediments rich in sulphate salts formed on Mars by evaporation from water.

The Odyssey orbiter has recently provided global maps of water-equivalent hydrogen that reveal surprisingly high abundances of near-surface hydrogen (~9–11 wt% H₂O equivalent⁴) in widespread equatorial regions where water ice is unstable to sublimation¹⁵. The observed H suggests the presence of hydrous silicates or salt hydrates in the upper 1 m of regolith^{16,17}. It has been known since Viking that lesser amounts of Cl and minor amounts of Br are associated with S, and the hydroxylated K–Fe³⁺ sulphate mineral jarosite (20% OH; 11% H₂O equivalent) has been found at Meridiani Planum¹, all of which point to a complex salt assemblage for which many possible mineralogies have been suggested¹⁸. Nevertheless, the widespread Mg–S association observed by Viking, Pathfinder and the MER rovers indicates that some form of MgSO₄·nH₂O is common in soils that are globally distributed on Mars.

The only common, naturally occurring members of the MgSO₄·nH₂O series on Earth are epsomite (MgSO₄·7H₂O, 51 wt% water), hexahydrite (MgSO₄·6H₂O, 47 wt% water) and kieserite (MgSO₄·H₂O, 13 wt% water). These three salts are believed to be the only members that occur on Earth as thermodynamically stable minerals¹⁹. Rare, metastable minerals of the series include pentahydrite (MgSO₄·5H₂O, 43 wt% water), starkeyite (MgSO₄·4H₂O, 37 wt% water) and sanderite (MgSO₄·2H₂O, 23 wt% water). Other hydration states ($n = 12, 3, 1.25$) are not recognized as minerals but can be synthesized. All of these salts consist of SO₄ tetrahedra and Mg(O, H₂O)₆ octahedra; some include extra-polyhedral water (water that is not in octahedral coordination with Mg)²⁰. Epsomite transforms readily to hexahydrite by loss of extra-polyhedral water; this transition is reversible and occurs at ~50–55% relative humidity (RH) at 298 K and at lower temperatures as the activity of water diminishes¹⁹. Kieserite is more stable at lower RH and higher temperature; for example, at moderate heating rates in thermogravimetric analysis the kieserite structure survives to ~670 K, compared with ~450 K for hexahydrite. However, as we show here, kieserite converts to hexahydrite or epsomite as humidity increases, yet these phases do not easily revert to kieserite on desiccation. Metastability, kinetic effects and pathway dependence are important factors in the MgSO₄·nH₂O system.

We have examined formation and transformation of MgSO₄·nH₂O minerals precipitated by evaporation from MgSO₄ solutions or transformed in the solid state under conditions of controlled temperature and humidity. These experiments include formation at conditions of ~1 torr total pressure, which is slightly less than Mars’ atmospheric pressure (~5 torr at most equatorial elevations). Other experiments have been conducted¹⁷ at temperatures from 100 to 298 K, and from 10^{–5} torr to ambient pressure (~580 torr in our laboratory at 2,265 m elevation). In addition, we have used thermogravimetric analysis (TGA) and isothermal controlled-humidity gravimetric analysis to examine water gain and loss from MgSO₄·nH₂O samples. Our most extensive data consist of X-ray diffraction (XRD) analyses collected while samples were held in an environmental cell²¹ at controlled RH and ambient temperature (~298 K) inside the diffractometer.

Results of these experiments show that the reaction path at very low RH ($\leq 0.5\%$) proceeds from precipitation of epsomite through transformation to hexahydrate, but the stable phase, kieserite, does not form. Instead, XRD analysis shows that hexahydrate in crystals of <1 mm becomes amorphous on timescales of a day (at 1 torr total pressure), a few hours (at 0.02 torr), or minutes (at $\sim 10^{-5}$ torr). This loss of crystal structure can also be observed at 298 K and 0.3–0.4% RH (~ 580 torr total pressure) on a timescale of 40 h (Fig. 1). An unanticipated observation was the retention of original epsomite crystal shapes, even as the crystal structure transformed to hexahydrate and then became amorphous. Desiccation does not lead to disintegration. The amorphous salt resists complete dehydration; after relatively rapid loss of two-thirds of the original 47% water in hexahydrate, the desiccation rate slows markedly. Even after desiccation at 1 torr and 0.5% RH for four months, amorphous salt prepared from centimetre-size hexahydrate crystals retains ~ 22 wt% water, 70% more than the amount of water in kieserite. Under colder conditions on Mars (~ 220 K equatorial average) and at typical Mars surface pressure (~ 5 torr) and diurnal RH ranging from $\sim 1\%$ to 100% (summer equatorial average $\sim 50\%$ RH²²), we expect that the rate of water loss and the kinetics of any transition of the amorphous phase to a crystalline form will be significantly reduced.

When kieserite rather than hexahydrate was exposed to comparable conditions of very low RH, no transition towards an amorphous salt was observed. This indicates that if the salts in soils or sediments on Mars contain kieserite, the kieserite could be stable at the currently dry martian surface but may actually contain less water than if the deposits were initially formed of epsomite/hexahydrate that has since become amorphous. Moreover, although kieserite resists dehydration in vacuum, experiments in the environmental cell show that it is easily hydrated on exposure to elevated humidity

($>55\%$ RH) where it converts to hexahydrate and then epsomite. When desiccated, these minerals become amorphous rather than transforming back to kieserite, indicating that kieserite might not be preserved in deposits that have experienced obliquity-driven cycles of hydration (water-ice stabilization) and desiccation as predicted for the upper 1–2 m of equatorial regolith on timescales of $\sim 10^6$ yr (ref. 15).

We have also examined amorphous $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$ on re-exposure to humid conditions. Figure 2 illustrates one of these experiments at 298 K and controlled RH of 31%. Within 460 h, amorphous material made from hexahydrate largely recrystallized to form hexahydrate, pentahydrate and starkeyite. Results from several environmental-cell desiccation and hydration experiments at 298 K and a range of RH conditions are summarized in Fig. 3. A significant implication of these experiments is that under conditions of water-ice stability near the martian equator, amorphous $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$ with diminished water content will not only rehydrate but may recrystallize to minerals that differ from the starting material. Under repeated changes in obliquity and equatorial climate, at the low average surface temperature of 220 K and average RH of $\sim 50\%$ shown in the lower part of Fig. 3, it is difficult to predict what the terminal salt mineral assemblage will be.

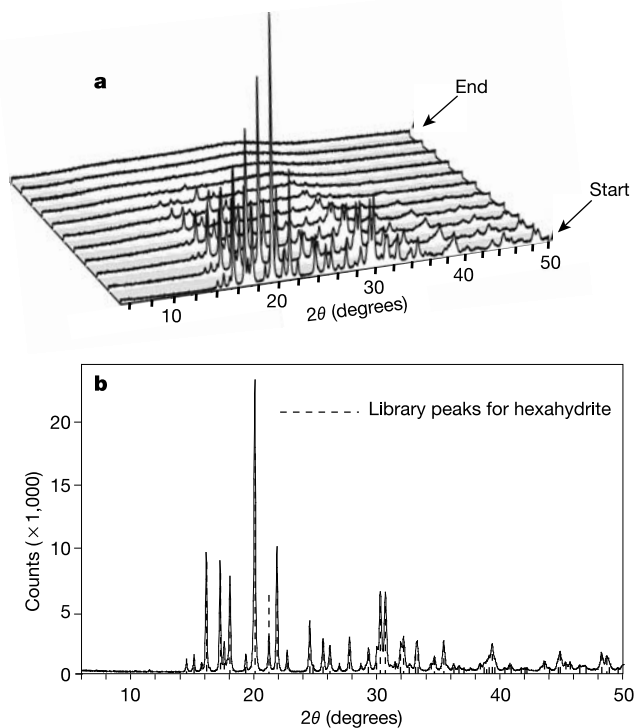


Figure 1 Dehydration experiment in an environmental cell. **a**, Ten sequential X-ray diffraction (XRD) patterns of hexahydrate desiccated at 298 K and 0.3–0.4% RH to produce amorphous $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$. The ten XRD patterns represent desiccation over a period of 40 h. **b**, The XRD pattern of the hexahydrate starting material. In Figs 1 and 2, XRD was performed using $\text{CuK}\alpha$ radiation.

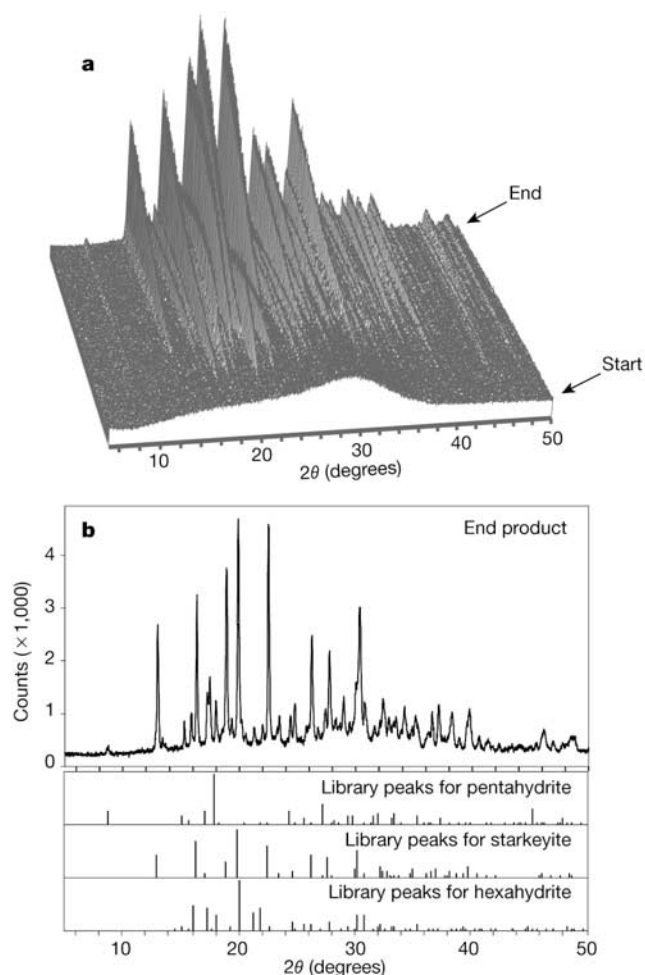


Figure 2 Rehydration experiment in an environmental cell. **a**, 115 sequential XRD patterns of amorphous $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$ rehydrated at 298 K and 31% RH over 460 h to form hexahydrate, pentahydrate and starkeyite; there is still some remnant amorphous material at 460 h. **b**, The XRD pattern of the partially recrystallized salts at the end of the experiment, with the library peak positions for the three minerals identified in the sample.

Nevertheless, it is likely that both bedded and soil-cementing salts carry a record of depositional and environmental parameters and thus have the potential to reveal geologic processes on Mars. As just one example, near-surface occurrences of $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$ as soil cement (10 wt% MgSO_4 on an anhydrous basis)^{7,8} are likely to account for variable portions of the water contents mapped by the Odyssey spacecraft at low latitudes where water ice is not at present stable⁵. The mineral or amorphous forms of $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$ in martian soils may vary with timing, duration, temperature and geographic distribution of water ice deposits during episodes of greater obliquity. If forms of $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$ also occur as part of the H-bearing mineral assemblage in evaporite sediments, the differences in salt-hydrate mineralogy between sediments and soils will reflect their differing modes of formation and subsequent exposure history.

The data summarized here suggest that studies of martian salts can provide vital information about the hydrogeologic history of Mars. Return of samples to Earth for detailed study will be important for reaching this goal. However, our results also show that mineralogy should be accurately characterized *in situ* before samples are removed from Mars. The ease with which dehydration/rehydration transformations take place in the $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$ system indicates that these and other environmentally sensitive salts collected on Mars probably will not be returned to Earth

unmodified unless exceptional effort is made to preserve martian conditions during sample return and recovery.

Examples from meteorites and the Moon illustrate this problem. Epsomite is observed as a hydrous phase in primitive CI1 chondrites and has been cited as evidence for late-stage oxidation of the CI1 parent body, but a strong case has been made that the epsomite formed after these meteorites were placed in humid terrestrial museums²³. Also, putative occurrences of goethite ($\text{FeO}(\text{OH})$) 'rust' in Apollo 16 rocks have been shown to have formed by hydration-oxidation of primary lawrencite (FeCl_2) (ref. 24) after removal from the Moon. This reaction occurred despite the short travel time and careful protection protocols (for example, storage in dry nitrogen after recovery) for lunar samples. What will we make of martian samples carried on a longer return if they contain unstable salt mixtures about which little is known? Accurate *in situ* mineralogical and chemical analysis is imperative if we are to reach the right conclusions from studies of saline, hydrated martian soils or sediments returned to Earth. □

Received 22 April; accepted 25 August 2004; doi:10.1038/nature02973.

- MER Rover web site. (<http://www.jpl.nasa.gov/mer2004/rover-images/mar-02-2004/images-3-2-04.html>) (March 2004).
- Brueckner, J. Determination of chemical composition of soils and rocks at the MER landing sites Gusev crater and Meridiani Planum using the APXS. *Eos* **85**(17, It. Assem. Suppl.), abstract V11A-05 (2004).
- Clark, B. C. Implications of abundant hygroscopic minerals in the martian regolith. *Icarus* **34**, 645–665 (1978).
- Feldman, W. C. *et al.* The global distribution of near-surface hydrogen on Mars. *J. Geophys. Res.* **109**, doi:10.1029/2003JE002160 (2004).
- Feldman, W. C. *et al.* Hydrated states of MgSO_4 at equatorial latitudes on Mars. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020181 (2004).
- Clark, B. C. *et al.* Inorganic analyses of martian surface samples at the Viking landing sites. *Science* **194**, 1283–1288 (1976).
- Toulmin, P. III *et al.* Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *J. Geophys. Res.* **82**, 4625–4634 (1977).
- Wänke, H., Brückner, J., Dreibus, G., Rieder, R. & Ryabchikov, I. Chemical composition of rocks and soils at the Pathfinder site. *Space Sci. Rev.* **96**, 317–330 (2001).
- Foley, C. N., Economou, T. & Clayton, R. N. Final chemical results from the Mars Pathfinder alpha proton X-ray spectrometer. *J. Geophys. Res.* **108**, doi:10.1029/2002JE002019 (2003).
- McSweeney, H. Y. *et al.* Basaltic rocks analyzed by the Spirit Rover in Gusev Crater. *Science* **305**, 842–845 (2004).
- Christensen, P. R. Martian dust mantling and surface composition: Interpretation of thermophysical properties. *J. Geophys. Res.* **87**, 9985–9998 (1982).
- Gendrin, A. & Mustard, J. F. Sulfate-cemented soils detected in TES data through the application of an automatic band detection algorithm. *Lunar Planet. Sci.* **XXXV**, abstract 1205 (2004).
- Clark, B. C. & Baird, A. K. Is the martian lithosphere sulphur rich? *J. Geophys. Res.* **84**, 8395–8403 (1979).
- Banin, A., Han, F. X. & Cicelsky, A. Acidic volatiles and the Mars soil. *J. Geophys. Res.* **102**, 13341–13356 (1997).
- Mellon, M. T. & Jakosky, B. M. The distribution and behavior of martian ground ice during past and present epochs. *J. Geophys. Res.* **100**, 11781–11799 (1995).
- Bish, D. L., Carey, J. W., Vaniman, D. T. & Chipera, S. J. Stability of hydrous minerals on the martian surface. *Icarus* **164**, 96–103 (2003).
- Bish, D. L., Vaniman, D. T., Fialips, C., Carey, J. W. & Feldman, W. C. Can hydrous minerals account for the observed mid-latitude water on Mars? *Sixth Int. Conf. on Mars*, abstract 3066 (2003).
- Clark, B. C. & Van Hart, D. C. The salts of Mars. *Icarus* **45**, 370–378 (1981).
- Chou, I.-M. & Seal, R. II Determination of epsomite-hexahydrite equilibria by the humidity-buffer technique at 0.1 MPa with implications for phase equilibria in the system $\text{MgSO}_4\text{--H}_2\text{O}$. *Astrobiology* **3**, 619–630 (2003).
- Hawthorne, F. C., Krivovichev, S. V. & Burns, P. C. in *Sulfate Minerals: Crystallography, Geochemistry, and Environmental Significance* (eds Alpers, C. N., Jambor, J. L. & Nordstrom, D. K.) 1–112 (Min. Soc. Am., Washington DC, 2000).
- Chipera, S. J., Carey, J. W. & Bish, D. L. Controlled-humidity XRD analyses: Application to the study of smectite expansion/contraction. *Adv. X-ray Anal.* **39**, 713–722 (1997).
- Savijärvi, H. Mars boundary layer modeling: Diurnal moisture cycle and soil properties at the Viking Lander 1 site. *Icarus* **117**, 120–127 (1995).
- Gounelle, M. & Zolensky, M. E. A terrestrial origin for sulfate veins in CI1 chondrites. *Meteorit. Planet. Sci.* **36**, 1321–1329 (2001).
- Taylor, L. A., Mao, H. K. & Bell, P. M. "Rust" in the Apollo 16 rocks. *Proc. 4th Lunar Sci. Conf.* 829–839 (1973).

Acknowledgements This research was supported by Los Alamos National Laboratory – Directed Research and Development Funding and by a NASA Mars Fundamental Research Program grant. Comments and suggestions by J. F. Bell and B. C. Clark helped to improve this Letter.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.V. (vaniman@lanl.gov).

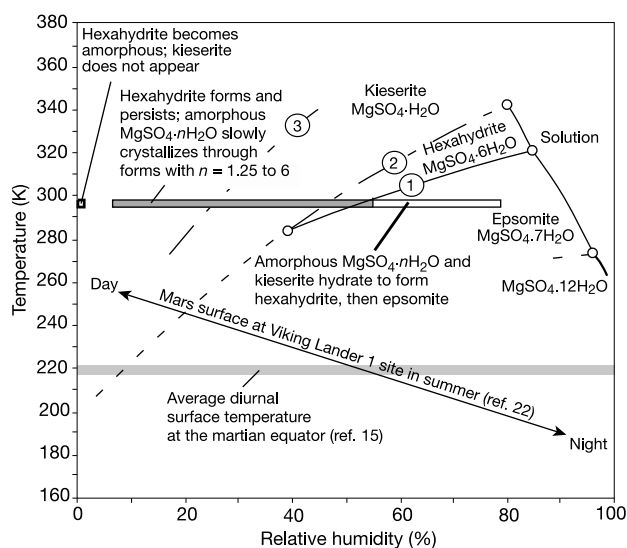


Figure 3 Results of controlled-humidity XRD experiments at 298 K plotted against stability fields for epsomite, hexahydrite and kieserite, modified from ref. 19. Their study provides accurate constraints on the epsomite to hexahydrite transition (curve 1); their estimate of the hexahydrite to kieserite transition (curve 2) is based on thermodynamic data extrapolated from the experimentally determined solution equilibrium; an alternative estimate of this transition is based solely on thermodynamic data (curve 3). Estimated stability of hexahydrite under martian conditions depends on extrapolation of curves 2 or 3. Stability of the $\text{MgSO}_4 \cdot 12\text{H}_2\text{O}$ phase is poorly constrained. Our experiments show that at RH of ~0.5%, hexahydrite forms from solution but becomes amorphous. Hexahydrite forms and persists at RH values from ~55% down to at least 7%. Amorphous $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$ formed at RH of ~0.5%, and subsequently exposed to RH of 7–55% crystallizes to various hydrates ($n = 1.25$ to 6) dependent on RH; ongoing long-term experiments suggest that crystalline end products vary with RH within the upper shaded region. Above RH of ~55%, both amorphous $\text{MgSO}_4 \cdot n\text{H}_2\text{O}$ and kieserite transform to hexahydrite and then epsomite. At martian conditions below 280 K, the rates and ranges for comparable reactions are not yet determined. The diurnal temperature–RH range shown (lower shaded region) for the Mars surface in summer at the Viking 1 site is from ref. 22. The average diurnal surface temperature of 220 K near the equator varies little despite changes in obliquity¹⁵.

A superconductor to superfluid phase transition in liquid metallic hydrogen

Egor Babaev^{1,2}, Asle Sudbø² & N. W. Ashcroft¹

¹Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York, 14853-2501, USA

²Department of Physics, Norwegian University of Science and Technology, N-7491 Trondheim, Norway

Although hydrogen is the simplest of atoms, it does not form the simplest of solids or liquids. Quantum effects in these phases are considerable (a consequence of the light proton mass) and they have a demonstrable and often puzzling influence on many physical properties¹, including spatial order. To date, the structure of dense hydrogen remains experimentally elusive². Recent studies of the melting curve of hydrogen^{3,4} indicate that at high (but experimentally accessible) pressures, compressed hydrogen will adopt a liquid state, even at low temperatures. In reaching this phase, hydrogen is also projected to pass through an insulator-to-metal transition. This raises the possibility of new state of matter: a near ground-state liquid metal, and its ordered states in the quantum domain. Ordered quantum fluids are traditionally categorized as superconductors or superfluids; these respective systems feature dissipationless electrical currents or mass flow. Here we report a topological analysis of the projected phase of liquid metallic hydrogen, finding that it may represent a new type of ordered quantum fluid. Specifically, we show that liquid metallic hydrogen cannot be categorized exclusively as a superconductor or superfluid. We predict that, in the presence of a magnetic field, liquid metallic hydrogen will exhibit several phase transitions to ordered states, ranging from superconductors to superfluids.

Hydrogen constitutes more than 90% of all atoms in the visible Universe and contributes three-quarters of its mass. It is widely accepted that hydrogen is abundant in the interiors of Saturn and Jupiter where it is both liquid and metallic⁵, and the origin of their magnetospheres. The conditions in these planets, particularly those of elevated temperatures, impel a view of dense hydrogen as a classical liquid metal⁶. In what follows, a quite different view is taken for low temperatures where, for a range of densities, hydrogen is projected to take up a state which may be described as a quantum liquid metal. The notion originates both with the light mass of the proton and the form of the electronically screened, and hence density dependent, proton–proton interactions.

The proton has one-fourth the mass of ⁴He, which in a condensed phase at normal conditions is a classic permanent liquid, a consequence of high zero-point energy compared with relatively weak ordering energies arising from interactions. Similarly, zero-point energies of protons in a dense environment are also high, and at elevated compressions there is a shift of electron density from intra-molecular regions to inter-molecular, and with it a progressive decline in the effective inter-proton attractions (both within proton pairs, and between). Because of this there is also a decline of ordering energies from interactions relative to protonic zero-point energies, and arguments have therefore been advanced² first to suggest the occurrence of a melting point maximum in compressed hydrogen, but second that there may also be a range of densities where, as in ⁴He, hydrogen may choose a fluid phase for its ground state. En route it passes through an insulator–metal transition and the phase could aptly be described as liquid metallic hydrogen (LMH), a translationally invariant two-component fermionic liquid. There is preliminary experimental evidence that a melting point maximum may indeed exist³ and it has received

recent theoretical backing⁴. Experimentally a 12.4-fold compression of hydrogen has already been achieved at around 320 GPa. Estimates suggest that LMH should appear at 13.6-fold compression at pressure in the vicinity of 400 GPa (ref. 4), whereas hydrogen alloys may exhibit metallicity at significantly lower pressures⁷. A predicted key feature of LMH at low temperature is the coexistence of superconductivity of proton–proton and electron–electron Cooper pairs⁸. These condensates are independently conserved, as electronic Cooper pairs cannot be converted to protonic Cooper pairs. Thus, there is no intrinsic Josephson coupling between the two condensates. This sets LMH apart from multi-component electronic condensates such as MgB₂. We therefore address some possible novel and experimentally observable physics of this new state of matter. Our goal is to discuss effects independent of pairing mechanism or other microscopic details. So in a search for qualitatively new physics we base our analysis solely on the topology of the proton–electron superconducting condensate.

The free energy appropriate for LMH will be described by the following Ginzburg–Landau (GL) model:

$$F = \frac{|\nabla + ie\mathbf{A}|\Psi_e|^2}{2m_e} + \frac{|\nabla - ie\mathbf{A}|\Psi_p|^2}{2m_p} + V(|\Psi_{e,p}|^2) + \frac{\mathbf{B}^2}{2}; \quad \mathbf{B} = \nabla \times \mathbf{A} \quad (1)$$

The condensate order parameters are complex fields denoted by $\Psi_\alpha = |\Psi_\alpha|e^{i\phi_\alpha}$, where $\alpha = p, e$, with p and e referring to protonic and electronic Cooper pairs and $V(|\Psi_\alpha|^2) = b_\alpha|\Psi_\alpha|^2 + (c_\alpha/2)|\Psi_\alpha|^4 + d|\Psi_p|^2|\Psi_e|^2$. We have introduced the masses m_e and m_p of the electronic and protonic Cooper pairs, respectively, and $\pm e$ is the effective charge of the Cooper pairs in the two condensates. Apart from the Josephson term $\Psi_e^*\Psi_p + h.c.$ which is forbidden, as noted above, equation (1) may include other terms which merely introduce small quantitative changes to the effects discussed in this paper, and which thus may be omitted. The GL free energy can be rewritten^{9,10} as:

$$F = \frac{1}{2} \frac{|\Psi_e|^2|\Psi_p|^2}{\frac{m_e}{2} + \frac{m_p}{2}} (\nabla(\phi_e + \phi_p))^2 + \frac{2}{\frac{|\Psi_e|^2}{m_e} + \frac{|\Psi_p|^2}{m_p}} \left\{ \frac{|\Psi_e|^2}{2m_e} \nabla \phi_e - \frac{|\Psi_p|^2}{2m_p} \nabla \phi_p - e \frac{|\Psi_e|^2}{m_e} + \frac{|\Psi_p|^2}{m_p} \right\} \left(\frac{2}{2} \right) + \frac{\mathbf{B}^2}{2} \quad (2)$$

The first term is recognized as the kinetic term of the Gross–Pitaevskii functional for liquid ⁴He, as no coupling to a vector potential $\mathbf{A}$ is present. This term may be thought of as describing an electrically neutral mode in the system, and is nothing but dissipationless co-directed currents of electrons and protons carrying zero net charge. The second term is equivalent to a gauge-invariant gradient term in an ordinary superconductor describing a charged mode in the system.

From the point of view of electronic pairing, it has been suggested that at certain densities metallic hydrogen is a type-II superconductor^{11,12}, that is, magnetic flux may penetrate it in the form of vortices. When both protonic and electronic pairings occur, the interesting physical question centres on whether there is a vortex structure for both, and there are several distinct possibilities. The main features of the ground state of vortex matter in this model are: (1) if the vortices of both components share the same core, such a composite vortex is characterized by phase windings ($\Delta\phi_e = \pm 2\pi$, $\Delta\phi_p = \mp 2\pi$), and then it has a finite energy per

unit length, and carries one flux quantum¹⁰. Only these types of composite vortices can actually be induced by a magnetic field. By a phase winding, we mean here the line integral of a phase gradient around a closed contour. In contrast, the vortices ($\Delta\phi_e = \pm 2\pi$, $\Delta\phi_p = 0$) and ($\Delta\phi_e = 0$, $\Delta\phi_p = \pm 2\pi$) carry a fraction of flux quantum and then have a logarithmically divergent energy per unit length¹⁰. (2) In the absence of an external magnetic field, the phase transitions in equation (1) are driven by a proliferation of thermally excited closed loops of vortices ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = 0$) and ($\Delta\phi_e = 0$, $\Delta\phi_p = 2\pi$) at critical temperatures T_c^e and T_c^p , respectively^{13,14}. We stress that in zero applied field the system is superconducting at all temperatures below T_c^e .

Next, we point out that application of an external magnetic field can change the physical state of LMH dramatically, and may result in a novel type of quantum fluid. We first consider the type-II regime. We emphasize that LMH should allow great flexibility in changing the GL parameter κ both for protons and electrons by varying the applied pressure and temperature^{11,12}. In a superconductor with only one type of Cooper pair, a lattice of Abrikosov vortices melts in a first-order phase transition at a temperature $T_M(B)$ which decreases with increasing magnetic field¹³. The physical possibilities for LMH are far richer, as we shall see. At zero temperature in an external field, the system allows only composite vortices with $\phi_p = -\phi_e$ (for such a vortex the first term in equation (2) is zero). However, because of thermal fluctuations, at $T \neq 0$ a vortex ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = -2\pi$) can split locally into two elementary vortices ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = 0$) + ($\Delta\phi_e = 0$, $\Delta\phi_p = -2\pi$), as shown in Fig. 1.

Such a splitting would result in a non-trivial contribution to the Ginzburg–Landau energy from the first term in equation (2), in the area in between two branches, because segments of such a loop attract each other logarithmically^{10,14}. The system defined by equation (1) therefore possesses a characteristic ‘vortex ionization’ temperature at which a composite flux line ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = -2\pi$) completely splits into two elementary vortices ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = 0$) + ($\Delta\phi_e = 0$, $\Delta\phi_p = -2\pi$). Such a splitting leads to a ‘plasma’ of line vortices interacting with a logarithmic potential. This topological transition is in the three-dimensional XY universality class, and should not be confused with topological phase transitions in two-dimensional superconductors. In Fig. 1, the protonic and electronic vortices are represented by thin and thick lines respectively. Of the two, the protonic vortex fluctuates more because it has a smaller stiffness $|\Psi_p|^2/m_p$ due to larger mass $m_p \gg m_e$. It carries a smaller fraction of flux quantum $\Phi = \Phi_0 |\Psi_p|^2/m_p [|\Psi_p|^2/m_p + |\Psi_e|^2/m_e]^{-1}$ (ref. 10) and thus has smaller energy per unit length. This leads to a ‘role inversion’ in vortex matter: vortices in the condensate of heavier particles cost less energy per unit length than vortices of condensates of lighter

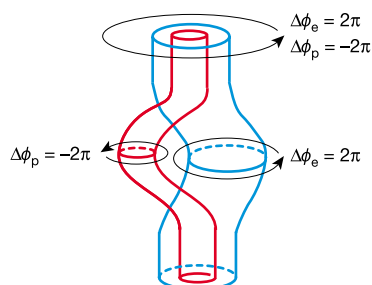


Figure 1 Local splitting of a composite vortex line in liquid metallic hydrogen, LMH. The blue and red colours represent electronic and protonic vortices, respectively. The length scales are chosen to be almost equal for graphical convenience.

particles. At low temperatures, the core size of a protonic vortex is expected to be much smaller than that of an electronic vortex because of the much larger mass of the former. However, such a picture is not applicable in the immediate vicinity of critical temperature for protons, where the protonic coherence length diverges. We now proceed to discuss the LMH phase diagram at low and high magnetic fields.

Let us first consider the low-magnetic field regime in Fig. 2, when the characteristic temperature required to split a composite vortex line, T_{SLM} , is much smaller than the melting temperature of the lattice of electronic vortices, T_M^e . Such a regime should be realizable in external fields much smaller than the upper critical magnetic field H_{c2}^e for the electronic condensate. Here, when the splitting of the field-induced composite vortex line ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = -2\pi$) → ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = 0$) + ($\Delta\phi_e = 0$, $\Delta\phi_p = -2\pi$) becomes of the order of intervortex distances, the logarithmic interaction of the split vortices would be screened in a manner similar to that expected in an ensemble of positively and negatively charged strings. When $T < T_{SLM} \ll T_M^e$ we therefore have an Abrikosov lattice of composite vortices, which we may call a superconducting superfluid because of the coexistence of a neutral and charged modes, denoted SSF in Fig. 2. However, upon transition to the ‘vortex-ionized’ state, $T > T_{SLM}$, we have a lattice of electronic vortices immersed in a liquid of protonic vortex lines. There has been a vortex sublattice melting; protonic superconductivity and the composite neutral mode disappear in this state, but the system remains in an electronic superconducting state so long as the electronic vortex lattice remains intact, that is, for $T < T_M^e$ (ref. 13). This phase is denoted ESC in Fig. 2; the phase transition separating SSF from ESC, as well

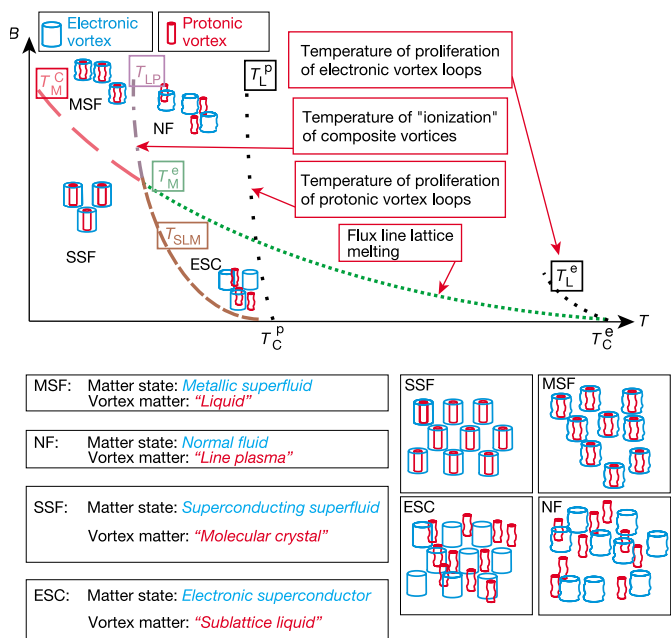


Figure 2 A schematic phase diagram of LMH as a function of applied magnetic field B and temperature T . Phase SSF: composite vortex lattice, which is a superconducting superfluid state. Phase MSF: composite vortex liquid, which is a non-superconducting metallic superfluid state. The transition from SSF to MSF is a superconductor–superfluid transition, and distinguishes LMH from any other known quantum fluid. Phase ESC: electronic vortex lattice immersed in a protonic vortex line liquid. This is a superconducting, but not superfluid, state. Phase NF: vortex line plasma, which emerges when composite vortex lines are fully ‘ionized’ into electronic as well as protonic vortex lines, neither arranged in a lattice. It features non-zero resistivity as well as viscosity.

as the phase ESC itself, have no counterparts so far in ordinary superconductors. One of the consequences of the presence of a background protonic vortex liquid is that electronic vortices carry only a fraction of the flux quantum, given¹⁰ by $\Phi = \Phi_0 |\Psi_e(T)|^2 / m_e [|\Psi_e(T)|^2 / m_e + |\Psi_p(T)|^2 / m_p]^{-1}$, where Φ_0 is the flux quantum. This fraction will be temperature dependent, and with increasing temperature should reach the value Φ_0 when $|\Psi_p| = 0$. In addition to the characteristic temperatures T_{SLM} and T_{M}^c , the system possesses characteristic temperatures T_{L}^p and T_{L}^e of thermally driven proliferation of protonic and electronic vortex loops, respectively, where $T_{\text{L}}^{p,e} > T_{\text{SLM}}$. The zero-field limit of $T_{\text{L}}^{p,e}$ corresponds to the temperatures $T_{\text{c}}^{p,e}$ introduced below equation (2), see Fig. 2.

We now consider LMH in strong magnetic fields. Here the characteristic temperature required to split a composite vortex line, T_{LP} , is much larger than the melting temperature of the lattice of composite vortices, T_{M}^c . Such a situation occurs when (1) the bare phase-stiffness of the electronic condensate $|\Psi_e|^2 / m_e$ is suppressed by the external magnetic field down to being of the same order of magnitude as the protonic stiffness, and (2) the temperature of the melting of the lattice of composite vortices is significantly lower than protonic and electronic critical temperatures (for example, the electronic GL parameter κ should be large, which should be achievable through choice of density^{11,12}). The phase diagram then features the following hierarchy of characteristic temperatures. First, T_{M}^c , the melting temperature of the lattice of composite vortices. Second, $T_{\text{LP}} > T_{\text{M}}^c$, the 'vortex liquid' to 'vortex plasma' transition temperature associated with fluxline splitting ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = -2\pi$) $\rightarrow$ ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = 0$) + ($\Delta\phi_e = 0$, $\Delta\phi_p = -2\pi$). As noted, this transition has a three-dimensional XY universality class. We emphasize that this transition is very different from the sublattice melting transition considered above.

Next, we examine the physical consequences of this hierarchy of characteristic temperatures. At low temperatures, the magnetic properties are controlled solely by the charged mode, which is described by the second term in equation (2). That is, at magnetic fields below the $T_{\text{M}}^c(B)$ line, the system exhibits a phase which is a field-induced lattice of composite vortices ($\Delta\phi_e = 2\pi$, $\Delta\phi_p = -2\pi$) for which $\phi_p = -\phi_e$ and the first term in equation (2) is exactly zero. This corresponds to the superconducting superfluid discussed above. However, increasing the magnetic field to cross the $T_{\text{M}}^c(B)$ line now leads to a first-order melting transition from a lattice to a liquid of composite vortices. This transition is completely decoupled from the neutral superfluid mode, while superconductivity is destroyed. It is therefore a first-order phase transition from a superconductor to a superfluid. This distinguishes LMH from any other known quantum fluid. It naturally requires a revision of current classification schemes of quantum fluids into the two categories of superconductors and superfluids. Indeed, the metallic superfluid state, denoted MSF in Fig. 2, acquires all the attributes of superfluidity of neutral atoms like ^4He even though microscopically it originates in a liquid of charged Cooper pairs.

We note that SSF phase is characterized by off-diagonal long-range order (ODLRO) in both fields $\lim_{|\mathbf{r}-\mathbf{r}'| \rightarrow \infty} \langle \Psi_\alpha(\mathbf{r}) \Psi_\alpha^\dagger(\mathbf{r}') \rangle \neq 0$ for $\alpha = p, e$. In the MSF state, the phases of both fields are disordered and there is no ODLRO for superconducting order parameters $\lim_{|\mathbf{r}-\mathbf{r}'| \rightarrow \infty} \langle \Psi_\alpha(\mathbf{r}) \Psi_\alpha^\dagger(\mathbf{r}') \rangle = 0$. In contrast, the neutral mode retains ODLRO, manifested by the preserved order in the phase sum ($\phi_p + \phi_e$). From this follows a counterpart to the Onsager-Penrose criterion¹⁵ for metallic superfluidity: $\lim_{|\mathbf{r}-\mathbf{r}'| \rightarrow \infty} \langle \Psi_p(\mathbf{r}) \Psi_p^\dagger(\mathbf{r}') \rangle \neq 0$. Under such circumstances, the system is incapable of sustaining a dissipationless charge current, yet is capable of sustaining dissipationless mass-flow and consequently a vortex lattice induced by rotation, as is possible in superfluid ^4He . Rotation of a high-pressure diamond cell containing hydrogen could be performed in the presence of a cooling system, making

such a rotating superfluid state experimentally accessible in principle. An even more intriguing possibility appears in the case of a rotation of liquid metallic deuterium, as the deuteron has also spin degrees of freedom. Increasing the temperature further produces an 'ionization' of composite vortices. Eventually superfluidity also disappears, and we are left with a metallic normal fluid; this corresponds to the phase denoted by NF in Fig. 2.

These observations should be of importance in experimentally establishing that hydrogen may indeed take up a low-temperature liquid metallic state. Experimental probes of the states of systems confined to high-pressure diamond cells are limited, but nonetheless, application of external fields as well as the use of induction coils have already been successfully used to detect superconductivity at high pressures. The latter technique should also permit flux noise experiments.

Our main points may be summarized as follows. (1) The vortex matter in LMH is different from vortex matter in ordinary metallic superconductors. Our analysis shows that starting from a system of two types of fermions which form two distinct types of Cooper pairs, we arrive at what can be viewed as a 'dual condensed matter of vortices'. The vortices can be mapped onto a system of two types of charged strings which may be viewed as 'extended line particles' with 'reversed' roles—the electronic vortices play the role of heavy particles, and the protonic vortices are the light particles. Then, the Abrikosov lattice of composite vortices may be interpreted as a molecular crystalline state, which at strong external fields undergoes at T_{M}^c a transition into a 'molecular liquid' and at higher temperature a transition to a 'plasma' state. In contrast, at weak external fields we find a 'sublattice melting' transition an intermediate state of vortex matter which has a counterpart in classical condensed matter physics—for example, atomic sublattice melting in AgI. (2) Our analysis shows that an experimental realization of LMH would mean that we have at hand a system which exhibits a phase transition from a superconductor to a superfluid, or vice versa, driven by a magnetic field. This counterintuitive fact may require a revision of the standard classification scheme of quantum liquids into superconductors and superfluids. □

Received 22 June; accepted 2 August 2004; doi:10.1038/nature02910.

- Hemley, R. J. & Mao, H.-K. Dense molecular hydrogen; order, disorder and localization. *J. Non-Cryst. Solids* **205**, 282–289 (1996).
- Ashcroft, N. W. The hydrogen liquids. *J. Phys. A* **12**, A129–A137 (2000).
- Datchi, F., Loubeyre, P. & LeToullec, R. Extended and accurate determination of the melting curves of argon, helium, ice (H_2O), and hydrogen (H_2). *Phys. Rev. B* **61**, 6535–6546 (2000).
- Bonev, S. A., Schwegler, E., Ogitsu, T. & Galli, G. A quantum fluid of metallic hydrogen. *Nature* (submitted).
- Guillot, T. Probing the giant planets. *Phys. Today* **57**, 63–69 (2004).
- Stevenson, D. J., Ashcroft, N. W. *Phys. Rev. A* **9**, 782–789 (1974).
- Ashcroft, N. W. Hydrogen dominant metallic alloys: High temperature superconductors? *Phys. Rev. Lett.* **92**, 187002 (2004).
- Moulopoulos, K. & Ashcroft, N. W. Generalized Coulomb pairing in condensed state. *Phys. Rev. Lett.* **66**, 2915–2918 (1991).
- Babaev, E., Faddeev, L. D. & Niemi, A. J. Hidden symmetry and duality in a charged two-condensate Bose system. *Phys. Rev. B* **65**, 100512 (2002).
- Babaev, E. Vortices with fractional flux in two-gap superconductors and in extended Faddeev model. *Phys. Rev. Lett.* **89**, 067001 (2002).
- Jaffe, J. & Ashcroft, N. W. Superconductivity in liquid metallic hydrogen. *Phys. Rev. B* **23**, 6176–6179 (1981).
- Jaffe, J. & Ashcroft, N. W. Critical fields of liquid superconducting metallic hydrogen. *Phys. Rev. B* **27**, 5852–5855 (1983).
- Fosheim, K. & Sudbø, A. *Superconductivity: Physics and Applications* Ch. 8 (Wiley & Sons, New York, 2004).
- Smiseth, J., Smørgrav, E. & Sudbø, A. Critical properties of the N-color London model. *Phys. Rev. Lett.* **93**, 077002 (2004).
- Penrose, O. & Onsager, L. Bose-Einstein condensation and liquid helium. *Phys. Rev.* **104**, 576–584 (1956).

Acknowledgements This work was supported by the National Science Foundation, the Research Council of Norway, NANOMAT and by STINT and the Swedish Research Council.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.B. (eb235@cornell.edu).

A quantum fluid of metallic hydrogen suggested by first-principles calculations

Stanimir A. Bonev, Eric Schwegler, Tadashi Ogitsu & Giulia Galli

Lawrence Livermore National Laboratory, University of California, Livermore, California 94550, USA

It is generally assumed^{1–3} that solid hydrogen will transform into a metallic alkali-like crystal at sufficiently high pressure. However, some theoretical models^{4,5} have also suggested that compressed hydrogen may form an unusual two-component (protons and electrons) metallic fluid at low temperature, or possibly even a zero-temperature liquid ground state. The existence of these new states of matter is conditional on the presence of a maximum in the melting temperature versus pressure curve (the ‘melt line’). Previous measurements^{6–8} of the hydrogen melt line up to pressures of 44 GPa have led to controversial conclusions regarding the existence of this maximum. Here we report *ab initio* calculations that establish the melt line up to 200 GPa. We predict that subtle changes in the intermolecular interactions lead to a decline of the melt line above 90 GPa. The implication is that as solid molecular hydrogen is compressed, it transforms into a low-temperature quantum fluid before becoming a monatomic crystal. The emerging low-temperature phase diagram of hydrogen and its isotopes bears analogies with the familiar phases of ³He and ⁴He (the only known zero-temperature liquids), but the long-range Coulomb interactions and the large component mass ratio present in hydrogen would result in dramatically different properties⁹.

The possible existence of low-temperature liquid phases of compressed hydrogen has been rationalized with arguments based on the nature of effective pair interactions and of the quantum dynamics at high density, resulting in proton–proton correlations insufficient for the stabilization of a crystalline phase⁵. But so far there has been no conclusive evidence establishing whether hydrogen metallizes at low temperature as a solid (the more widely accepted view to date) or as a liquid. Measurements and theoretical predictions of the near-ground-state high-pressure phases³ of hydrogen are difficult because of the light atomic mass, significant quantum effects and strong electron–ion interactions. In this regard, the finite temperature liquid–solid phase boundary predicted here is especially valuable for understanding the manner in which hydrogen metallizes.

The appearance of a maximum melting temperature in hydrogen is in itself a manifestation of an unusual physical phenomenon. The few systems with a negative melt slope involve either open crystalline structures, such as water and graphite, or in the case of closed packed solids, a promotion of valence electrons to higher orbitals upon compression (6s to 5d in caesium¹⁰, for example). In these cases, the liquid is denser than the solid when they coexist, possibly because of structural or electronic transitions taking place continuously in the liquid, as a function of pressure, but only at discrete pressure intervals in the solid.

In contrast, recent experiments⁸ have shown that hydrogen phase I—a solid structure with rotationally free molecules associated with the sites of a hexagonal close packed (h.c.p.) lattice—persists below the liquid transition up to at least 150 GPa, that is, well beyond the melt curve maximum predicted here. The promotion of electrons to higher orbitals in hydrogen can also be ruled out because of the prohibitive high energy involved. Alternatively, it has been suggested that dissociation in the fluid, either gradual⁷ or following a first-order liquid–liquid phase transition^{11,12}, may be the origin of

a maximum in the melt line. This idea is not supported by our results. Instead, we explain the physical origin of the maximum in terms of changes in the intermolecular interactions—a mechanism significantly different from that expected from familiar phenomenological models.

The approach taken here to compute the melt line is one of direct simulation of the melting process. Hysteresis effects of superheating or super-cooling during the phase transition are avoided by simulating solid and liquid phases in coexistence. The validity of this method is then assessed by reducing size effects in such a way that realistic thermodynamic processes can be mimicked. This technique, known as two-phase simulation, has mostly been applied with model potentials¹³, thus allowing the simulation of large systems, and only recently was it demonstrated^{14,15} that implementations within the framework of first-principles molecular dynamics (MD) are feasible.

Two-phase simulations (Fig. 1) are performed for various pressure and temperature conditions to predict the melt curve of hydrogen up to 200 GPa (Fig. 2). As discussed in refs 7 and 8, the experimental data available up to 44 GPa can be equally well fitted with several empirical melt equations, leading to qualitatively different conclusions regarding the further rise or fall of the phase boundary. Our results at 50 GPa compare favourably with recent measurements⁸, and the calculations at 130 and 200 GPa predict a negative slope in this pressure region.

Previous first-principles calculations^{11,16} indicated the existence of a sharp transition from molecular to non-molecular fluid, thus raising the possibility that the melt curve maximum is a liquid–liquid–solid triple point. We therefore studied the properties of the liquid at temperatures above melting. In our simulations, we do find

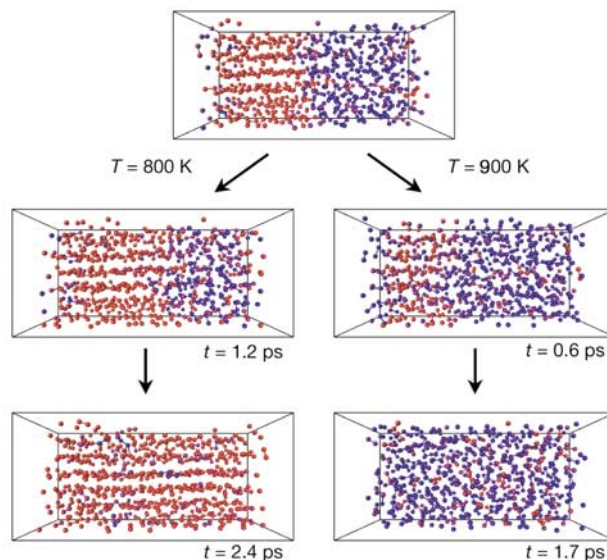


Figure 1 Snapshots from two-phase MD simulations at $P = 130$ GPa and temperatures below and above the melting temperature. A quantitative assessment of the instantaneous local order environment around each molecule has been performed as described in ref. 29, and compared with single-phase solid and liquid simulations. Molecules are coloured according to the arrangement of their nearest neighbours, red and blue representing configurations uniquely characteristic of the h.c.p. solid and liquid at the given P and T , respectively (the Q_4 order parameter²⁹ has been used for the colour map). The systems reach equilibrium at the target average temperatures (800 and 900 K) over a time interval of ~ 0.5 ps, which is followed by periods of coexistence lasting from one to several picoseconds. The phase transitions are observed by monitoring changes in the diffusion constants, pair correlation functions, specific volumes and local order parameters. In addition, some noticeable correlations between temperature and diffusion variations indicate the exchange of latent heat.

a first-order liquid–liquid phase transition from molecular to dissociating fluid, but this occurs at temperatures higher than those of the melting line; at 200 GPa, the liquid–liquid transition takes place between 900 and 1,000 K (Fig. 2). Thus, over the pressure range considered here, the liquid remains molecular below 900 K, with molecules stable over the simulation times of several picoseconds. We emphasize that on heating above a critical temperature, the transition to a non-molecular fluid takes place very rapidly, over a couple of hundred femtoseconds of simulation time. When the fluid is subsequently quenched, the molecules recombine and the system reverts to the molecular state. Therefore, we conclude that the change from a positive to a negative melt line slope happens gradually and we emphasize that it is not directly related to molecular dissociation. However, extrapolations of the melt line and the liquid–liquid phase transition indicate a triple point at ~ 300 GPa and 400 K. Above this pressure, the solid is expected to melt into a metallic liquid.

To estimate the errors originating from finite system size effects, we have computed the ground-state pressure and energy for different size simulation cells sampled at the Γ -point and with a $2 \times 2 \times 2$ k-point mesh. The pressure and energy differences between the solid and the liquid are converged to within ~ 0.3 GPa and ~ 1 meV per ion, and the estimated errors are included in Fig. 3. A question that remains to be answered is whether the size of the system considered here is sufficient to simulate liquid–solid coexistence. To address this, we carried out MD simulations of single-phase solid and liquid hydrogen at the melting temperatures determined in the two-phase simulations. The differences in the specific volumes and enthalpies obtained in this way (Fig. 3) are then used to compute the melt slopes using the Clausius–Clapeyron equation. The agreement with the results from two-phase simulations gives us confidence that the computed melt curve is thermodynamically consistent.

We also discuss the validity of the principal approximations in

our theoretical method: (1) adiabatic interactions, (2) the generalized gradient approximation (GGA) for the exchange–correlation energy, and (3) classical treatment of ion motion. Electron–phonon interactions are significant in the solid phase when the direct bandgap is comparable with the phonon frequencies, an event that eventually occurs at high densities. But at 200 GPa and 600 K—the conditions examined here where the electron–phonon coupling is expected to be most pronounced—the GGA bandgap is still ~ 2 eV. We note that the GGA tends to systematically underestimate bandgaps; this has been shown for various materials, including hydrogen¹⁷. However, because the highest vibrational frequencies in hydrogen are ~ 0.5 eV, the computed GGA bandgap—a lower bound to the actual value—is already sufficient to establish that non-adiabatic effects are indeed negligible.

Local density approximations are also expected to favour the phase with more delocalized electrons; previous studies have shown a consistent tendency of the GGA to underestimate melting temperatures, such as those of lithium hydride¹⁴ and aluminium¹⁵. Here, this effect is expected to be very small because the molecular bonding properties and the atomic coordination differ little in the solid and liquid phases. This also implies that the zero-point energy bound in the vibron motion is similar in the two phases. Indeed, velocity–velocity autocorrelation analysis carried out on our MD trajectories shows that vibron frequency distributions (containing all vibrational modes) differ by less than 10 cm^{-1} in the solid and the liquid between 50 and 200 GPa. In addition, we have computed the first-order quantum correction to the ionic free energies. It is given by the Wigner–Kirkwood formula, $\Delta F = \hbar^2 / (24 k_B T^2) \sum_i \langle \mathbf{F}_i^2 \rangle / m_i$, where the average is over the classical ensemble, and $\mathbf{F}_i$ and m_i are the ionic forces and masses. The difference in the quantum corrections for the phases obtained in this way at pressures of 50, 130 and 200 GPa remains less than 2 meV for temperatures near the computed melt curve.

We now turn our attention to the physical origin of the maximum in the melt curve of hydrogen. The transition from an ordered to a

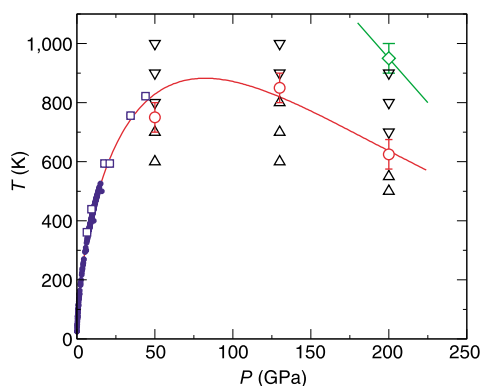


Figure 2 Melt curve of hydrogen predicted from first-principles MD. The filled circles are experimental data from refs 6 and 7 and references therein, and the open squares are measurements from ref. 8. Triangles indicate two-phase simulations where solidification (up) or melting (down) have been observed, and bracketed melting temperatures (T_m) are represented by open circles. As the phase boundary is approached, the period of coexistence increases and eventually the outcome becomes dependent on the choice of simulation parameters. This degree of arbitrariness is reflected in the error bars of T_m , which also include the standard deviation of the temperatures collected during the MD simulations. All experimental and theoretical points are given equal weight and fitted with a Kechin melt equation³⁰ (solid line in the figure): $T_m = 14.025(1 + P/a)^b \exp(-cP)$ K, where P is in units of GPa, $a = 0.030355$, $b = 0.59991$, and $c = 0.0072997$. The open diamond marks the liquid–liquid transition from molecular to non-molecular fluid at 200 GPa, and the estimated slope of this phase boundary is given by the green line. The error bar on the diamond symbol indicates the hysteresis effects during the simulation of the liquid–liquid transition.

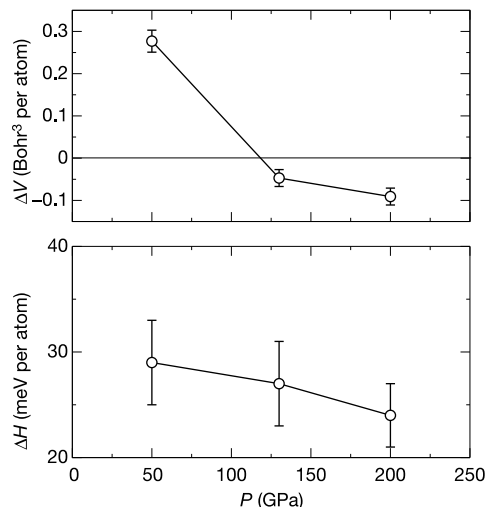


Figure 3 Difference of the specific volumes (ΔV) and enthalpies (ΔH) between the liquid and solid phases at the melting temperatures determined from the two-phase simulations. The reported uncertainties include standard deviations collected during the MD simulations and observed size effects from single-phase simulations with 360 and 768 atom supercells. The data are used to compute melt slopes from the Clausius–Clapeyron equation: $dT_m/dP = T_m \Delta V / \Delta H$; the values obtained for 50, 130 and 200 GPa are (6.5 ± 1.2) , (-1.4 ± 0.6) and (-2.3 ± 0.6) K/GPa, respectively. For comparison, the slopes from the melt line fit in Fig. 2 are 3.9, -2.2 and -2.7 K GPa $^{-1}$, but we note that these are weighted heavily by the experimental data, especially at lower pressures.

disordered state is governed by the relative importance of entropy and energy. At high pressure, the interactions become mostly repulsive, and it is reasonable to expect that a deviation from a crystalline arrangement becomes progressively more unfavourable in terms of enthalpy. This is indeed the case for most materials, as is reflected in the positive slopes of their melt curves. On the other hand, attractive many-body interactions can soften the steep increase of the repulsive forces at high density^{18,19}. Although, in principle, such a softening opens up the possibility for a melt curve maximum²⁰, it does not appear to be a sufficient condition for its existence. Indeed, softening of the repulsive part of the pair potentials has been observed for a number of materials for which there is no experimental indication of a melt line turnover. As will be shown below, the physics behind the unique melt curve of hydrogen is related to the different rate at which the repulsive interactions (at intermolecular range) are weakened in the solid and liquid phases as a function of pressure.

To investigate how many-body interactions give rise to repulsive force softening in the liquid and the solid, we have performed an analysis based on maximally localized Wannier functions (MLWF). These functions have been shown²¹ to provide an intuitive description of chemical bonds in condensed phases and they have recently been used²² to investigate the infrared activity of candidate structures of hydrogen phase III. We have computed MLWF along the 700 K isotherm, at various pressures, and find that their spreads (Fig. 4) increase faster in the liquid than in the solid, as the system is compressed. Importantly, the differences in the two phases come

from the tails of the MLWF, whereas their profiles around the centres of the molecular bonds remain nearly the same. Therefore, the large MLWF spread observed in the high-pressure fluid comes from the increased overlap of the molecular orbitals, here represented by the MLWF. Because the specific volumes of the solid and the liquid are very similar under the conditions examined here, the physical origin of the larger spread in the fluid can be attributed to the disorder in the liquid phase. Furthermore, the tails of the MLWF in the fluid are characterized by an asymmetry that can be viewed as an effective intermolecular charge transfer, giving rise to an enhanced infrared activity in the liquid phase.

The properties of the MLWF found in our calculations are indicative of the behaviour of repulsive interactions in the solid and liquid as a function of pressure. Effective hydrogen–hydrogen pair potentials derived from our *ab initio* MD trajectories using a force matching technique²³ are displayed in the inset of Fig. 4. As in the MLWF behaviour, we observe a stronger softening of the potential in the liquid than in the solid. Our Wannier function analysis indicates that this increased softening is the result of charge transfer processes, enhanced by disorder, which appear to have a qualitative character similar to those proposed in ref. 24 for phase III of solid hydrogen. We also note that the attenuation of repulsive interactions observed here in the solid phase correlates well with the analysis of Raman spectroscopic data²⁵, which has indicated softening of the effective intermolecular force constants above ~100 GPa.

To summarize, we have predicted the melt curve of hydrogen between 50 and 200 GPa from first principles. Above approximately 82 GPa, the melt line has a negative slope—a phenomenon that we have related to softening of the intermolecular interactions, which occur at a faster rate in the liquid than in the solid, as a function of pressure. Our results provide strong evidence that above ~300 GPa the solid melts into a metallic liquid and that a low-temperature metallic quantum fluid will exist at pressures near 400 GPa. Although we have only directly probed temperatures as low as 600 K, our conclusions are likely to hold at lower temperatures as well. Indeed, quantum ion motion usually favours phases with lower ion–ion correlations, which would further stabilize the low-temperature liquid phase found here. In addition, our findings are consistent with pressure estimates from experiments where hydrogen is expected to metallize^{3,26}. Finally, we note that the observed increase in MLWF spreads when going from the high-pressure solid to the molecular liquid phase suggests the presence of dynamically induced intermolecular charge transfer (dipole moments); this indicates that the transition from solid to liquid at high pressure is accompanied by an increase of the hydrogen infrared absorption. Therefore, we propose that measurements of infrared activity could be used to verify experimentally the predicted turnover of the melt curve. □

Methods

Molecular dynamics

Our *ab initio* MD simulations are carried out with the GP code, written by F. Gygi, and using the Car–Parrinello (CP) method²⁷; the many-body electron problem was solved quantum-mechanically within the generalized gradient approximation²⁸ of density functional theory, while the ions are propagated classically (the validity of this approximation for the problem at hand is discussed in the text). The actual calculations are performed for deuterium instead of hydrogen. The reason for this is that in the classical-ion picture, for the quantities of relevance here there is no distinction between hydrogen and its isotopes, but the substitution with heavier ions allows us to integrate the equations of motion with reasonably large time steps: 2 and 3 a.u. (1 a.u. = 0.0242 fs), depending on the density. Furthermore, we use a Γ -point sampling of the Brillouin zone, Troullier–Martins pseudopotentials, and a 60-Ry energy cut-off for the plane wave expansion of the Kohn–Sham orbitals. The accuracy and transferability of our pseudopotential has previously been extensively tested¹⁶. We have explicitly verified that the electrons remain near the Born–Oppenheimer surface throughout the simulations by comparing pressures determined in the CP runs with those determined by fully optimizing wavefunctions for the same ionic trajectories. The use of CP-MD results in a systematic shift of about 0.5 GPa in both the solid and liquid phases.

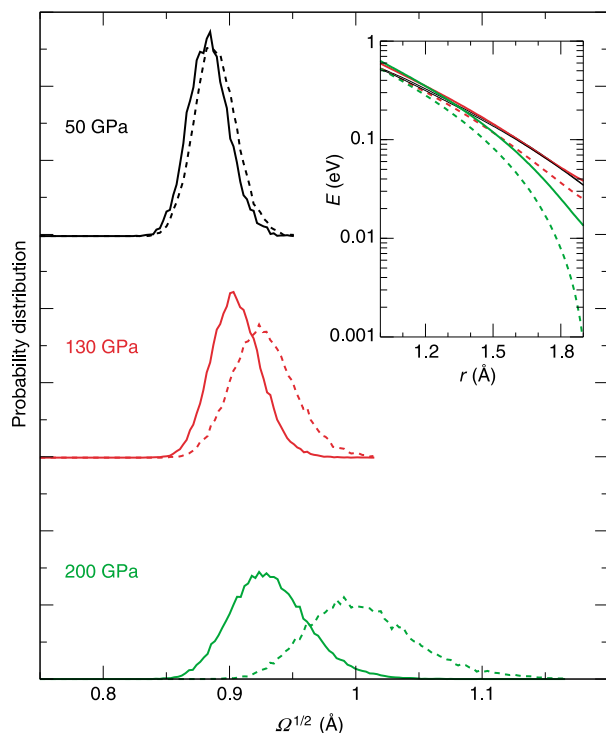


Figure 4 The probability distribution of MLWF spreads²¹ at $T = 700$ K and $P = 50$ (black curves), 130 (red curves) and 200 GPa (green curves). The solid and dashed curves correspond to distributions collected from the solid and liquid phases, respectively. The inset shows the effective hydrogen–hydrogen potentials obtained directly from the simulation trajectories with the force-matching procedure described in ref. 23 (the same colouring has been used). As the pressure is increased, the spreads of the MLWF are shown to increase more quickly in the liquid than in the solid phase. In addition, effective potentials soften more quickly in the liquid than in the solid phase at the nearest-neighbour molecular separation distance. Note that the potentials at 50 GPa in the liquid and solid phases, and at 130 GPa in the solid phase, are indistinguishable on this scale.

Two-phase simulations

We perform the two-phase simulations by equilibrating solid (in the h.c.p. structure) and liquid phases separately in 360-atom supercells with fixed geometry and applied periodic boundary conditions. The atomic arrangements obtained in this way are then merged, and the liquid and solid parts of the resulting 720-atom supercells are also heated independently to reduce the strain near the newly formed interface. The whole equilibration process typically lasts for several picoseconds of simulation time. The MD runs of coexisting phases are then performed in the N - P - T ensemble (constant number of particles, pressure and temperature respectively). The actual system being simulated initially represents semi-infinite alternating slabs (by applying periodic boundary conditions) of solid and liquid, but by the end of the runs only one of the two phases, the one that is stable at the chosen P and T , fills the entire volume.

Received 27 April; accepted 25 August 2004; doi:10.1038/nature02968.

- Wigner, E. & Huntington, H. B. On the possibility of metallic modifications of hydrogen. *J. Chem. Phys.* **3**, 764–770 (1935).
- Ashcroft, N. W. Metallic hydrogen: A high temperature superconductor? *Phys. Rev. Lett.* **21**, 1748–1799 (1968).
- Mao, H. K. & Hemley, R. J. Ultrahigh-pressure transitions in solid hydrogen. *Rev. Mod. Phys.* **66**, 671–692 (1994).
- Brovman, E. G., Kagan, Yu. & Kholas, A. Properties of metallic hydrogen under pressure. *Sov. Phys. JETP* **35**, 783–792 (1972).
- Ashcroft, N. W. The hydrogen liquids. *J. Phys. Condens. Matter* **12**, A129–A137 (2000).
- Diatschenko, V. *et al.* Melting curves of molecular hydrogen and molecular deuterium under high pressures between 20 and 373 K. *Phys. Rev. B* **32**, 381–389 (1985).
- Datchi, F., Loubeyre, P. & LeToullec, R. Extended and accurate determination of the melting curves of argon, helium, ice (H_2O) and hydrogen (H_2). *Phys. Rev. B* **61**, 6535–6546 (2000).
- Gregoryanz, E., Goncharov, A. F., Matsuishi, K., Mao, H.-k. & Hemley, R. J. Raman spectroscopy of hot dense hydrogen. *Phys. Rev. Lett.* **90**, 175701 (2003).
- Ashcroft, N. W. Hydrogen at high density. *J. Phys. A* **36**, 6137–6147 (2003).
- Jayaraman, J., Newton, R. C. & McDonough, J. M. Phase relations, resistivity, and electronic structure of cesium at high pressure. *Phys. Rev.* **159**, 527–533 (1967).
- Scandolo, S. Liquid–liquid phase transition in compressed hydrogen from first-principles simulations. *Proc. Natl Acad. Sci. USA* **100**, 3051–3053 (2003).
- Kechin, V. V. Melting of metallic hydrogen at high pressure. *JETP Lett.* **79**, 46–49 (2004).
- Morris, J. R., Wang, C. Z., Ho, K. M. & Chan, C. T. Melting line of aluminum from simulations of coexisting phases. *Phys. Rev. B* **49**, 3109–3115 (1994).
- Ogitsu, T., Schwegler, E., Gygi, F. & Galli, G. Melting of lithium hydride under pressure. *Phys. Rev. Lett.* **91**, 175502 (2003).
- Alfe, D. First-principles simulations of direct coexistence of solid and liquid aluminum. *Phys. Rev. B* **68**, 064423 (2003).
- Bonev, S. A., Militzer, B. & Galli, G. *Ab initio* simulations of dense liquid deuterium: Comparison with gas-gun shock-wave experiments. *Phys. Rev. B* **69**, 104101 (2004).
- Chacham, H., Zhu, X. & Louie, S. G. Pressure-induced insulator-metal transitions in solid xenon and hydrogen: A first-principles quasiparticle study. *Phys. Rev. B* **46**, 6688–6699 (1992).
- Hemley, R. J. *et al.* Equation of state of solid hydrogen and deuterium from single-crystal x-ray diffraction to 26.5 GPa. *Phys. Rev. B* **42**, 6458–6470 (1990).
- Nagao, K., Bonev, S. A., Bergara, A. & Ashcroft, N. W. Enhanced Friedel structure and proton pairing in dense solid hydrogen. *Phys. Rev. Lett.* **90**, 035501 (2003).
- Yoshida, T. & Kamakura, S. Liquid-solid transitions in systems of soft repulsive forces. *Prog. Theor. Phys.* **52**, 822–838 (1972).
- Marzari, N. & Vanderbilt, D. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* **56**, 12847–12856 (1997).
- Souza, I., Martin, R. M., Marzari, N., Zhao, X. & Vanderbilt, D. Wannier-functional description of the electronic polarization and infrared absorption of high-pressure hydrogen. *Phys. Rev. B* **62**, 15505–15519 (2000).
- Izvekov, S., Parrinello, M., Burnham, C. J. & Voth, G. A. Effective force fields for condensed phase systems from *ab initio* molecular dynamics simulations: A new method for force-matching. *J. Chem. Phys.* **120**, 10896–10913 (2004).
- Hemley, R. J., Soos, Z. G., Hanfland, M. & Mao, H. K. Charge-transfer states in dense hydrogen. *Nature* **369**, 384–387 (1994).
- Moshary, F., Chen, N. H. & Silvera, I. F. Pressure dependence of the vibron in H_2 , HD, and D_2 : Implications for inter- and intramolecular forces. *Phys. Rev. B* **48**, 12613–12619 (1993).
- Loubeyre, P., Ocelli, F. & LeToullec, R. Optical studies of solid hydrogen to 320 GPa and evidence for black hydrogen. *Nature* **416**, 613–617 (2002).
- Car, R. & Parrinello, M. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
- Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
- Steinhardt, P. J., Nelson, D. R. & Ronchetti, M. Bond-orientational order in liquids and glasses. *Phys. Rev. B* **28**, 784–805 (1983).
- Kechin, V. V. Melting curve equations at high temperature. *Phys. Rev. B* **65**, 052102 (2001).

Acknowledgements We thank S. Izvekov for providing the force-matching routines that were used in the analysis shown in the inset of Fig. 4, and F. Gygi for many discussions. This work was performed under the auspices of the US Department of Energy at the University of California/Lawrence Livermore National Laboratory.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.A.B. (bonev1@llnl.gov).

Room-temperature ferromagnetic nanotubes controlled by electron or hole doping

L. Krusin-Elbaum, D. M. Newns, H. Zeng, V. Derycke*, J. Z. Sun & R. Sandstrom

IBM Research Division, T. J. Watson Research Center, Yorktown Heights, New York 10598, USA

* Present address: Université de Paris-Sud XI/Orsay, Bât. 462, 91191 Gif-sur-Yvette, France

Nanotubes and nanowires with both elemental^{1,2} (carbon or silicon) and multi-element^{3–5} compositions (such as compound semiconductors or oxides), and exhibiting electronic properties ranging from metallic to semiconducting, are being extensively investigated for use in device structures designed to control electron charge^{6–8}. However, another important degree of freedom—electron spin, the control of which underlies the operation of ‘spintronic’ devices⁹—has been much less explored. This is probably due to the relative paucity of nanometre-scale ferromagnetic building blocks¹⁰ (in which electron spins are naturally aligned) from which spin-polarized electrons can be injected. Here we describe nanotubes of vanadium oxide (VO_x), formed by controllable self-assembly¹¹, that are ferromagnetic at room temperature. The as-formed nanotubes are transformed from spin-frustrated semiconductors to ferromagnets by doping with either electrons or holes, potentially offering a route to spin control¹² in nanotube-based heterostructures¹³.

Electrons in the 3d transition metals (TM) and their oxides¹⁴ combine the properties of interactions strong enough to enable magnetism with adequate mobility to permit effective spin transport. However, only the transition-metal oxides, (TM) O_x , owing to the variety of ways in which (TM) O_4 tetrahedra and (TM) O_6 octahedra can be edge- and corner-linked into low-dimensional (layered) structures, are sufficiently rich in structural possibilities for self-assembly of novel functionalities at the nanoscale. Two well-known examples of functionalities arising from strong electron correlations are high-temperature superconductivity in copper oxides—containing CuO_2 planes which are transformed from insulating (antiferro) magnets into superconductors by doping with charge—and colossal magnetoresistance in manganites¹⁵. Bringing such phenomena down to the nanometre scale is a challenge, and a recent synthesis¹⁶ of hollow (~800 nm in diameter with aspect ratio 1:5) cylinders of a manganite $La_{0.325}Pr_{0.3}Ca_{0.375}MnO_3$ is an encouraging step toward spin-polarized transport using nanotubes.

Among the transition-metal oxides, the oxides of vanadium are exceptionally rich in both structural diversity¹⁷ and electronic properties¹⁸, with the potential to be assembled into spin-tunable nanoforms¹⁹. Vanadium oxides (VO_x , with $1.0 < x < 2.5$) are frequently of mixed valency (typically V^{4+} and V^{5+}), often crystallizing into weakly coupled, layered two-dimensional (2D) networks. Owing to the interplay of internal degrees of freedom—spin, charge and orbital—they can suddenly transform from a metal to a (Mott–Hubbard) insulator^{15,18,20}, become superconducting²¹, or host unusual quantum spin states^{22–24}. Our focus is on the unexplored spin textures of mixed-valency VO_x multi-wall nanotubes that are self-assembled using dodecylamine¹¹ ($C_{12}H_{2n+1}NH_2$, where $n = 12$) as a structure-directing surfactant molecule (Fig. 1a–c). We find that in their as-assembled state, $S = 1/2$ spins on the V^{4+} sites form a (quantum) spin-liquid^{15,25} with a spin-gap $\Delta_s \approx 665$ K. A distinct feature of Mott insulators is that doping (that is, changing the particle density) can greatly affect their properties. We discover that the tubes, when $\pm$ charge-doped, develop a nonlinear (ferro-

magnetic) response to the applied magnetic field. We discuss how this ferromagnetism, supported by particle-hole complementarity, can arise from the low dimensionality of the nanotube structure with crystal field splittings that enforce a Mott gap.

To set the stage for the *ex situ* doping experiments, we first discuss the structure of VO_x nanotube walls. The elemental composition after hydrothermal synthesis¹¹ is VO_x[C₁₂H₂₈N]_{0.27}, with $x \approx 2.4$. Structural information^{11,26} derived from the X-ray powder diffraction patterns (as in Fig. 1d) indicates that the *d* spacing in the 00*l* direction is controlled by the dodecylammonium chains, whereas the VO_x wall structure, with the tetragonal basal plane parameter $a = 6.144 \text{ \AA}$, is closely related^{11,26} to that of BaV₇O₁₆·*n*H₂O (ref. 27), in which six of the vanadium atoms are in octahedral coordination and the seventh is in a tetrahedrally coordinated site. On the basis of this structure, the nanotube wall is double-layered, the bottom layer

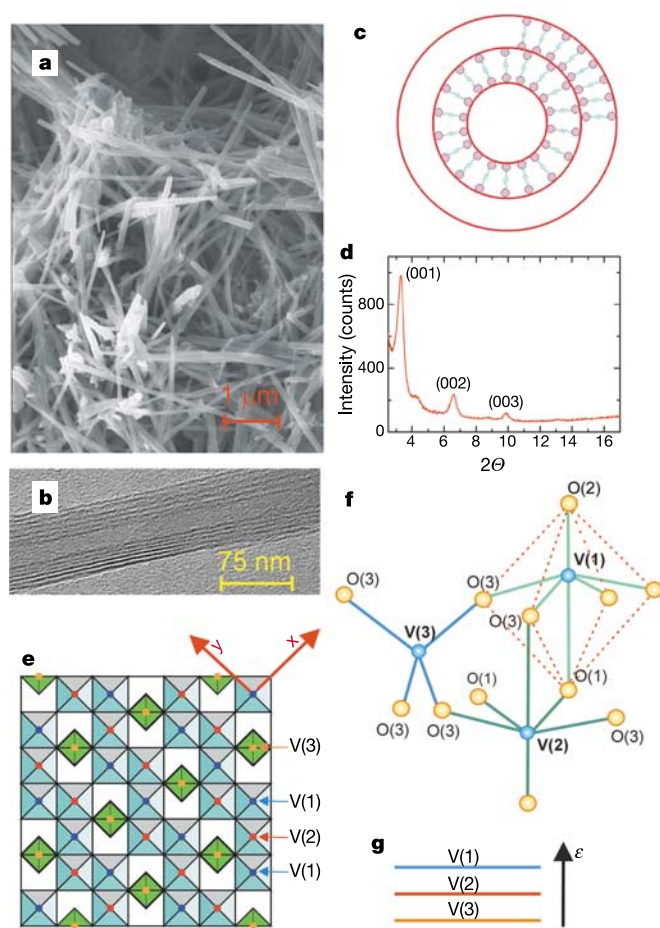


Figure 1 Vanadium oxide nanotubes ‘self-assembled’ using dodecylamine as a structure-directing templating agent. **a**, Scanning electron microscopy (SEM) images show that the tubes (typically open-ended) can be several micrometres long, with diameters that vary from 60 to 100 nm. A transmission electron microscopy (TEM) image in **b** shows the tubes to be multi-walled (they appear as dark fringes owing to high scattering of VO_x layers). Between these vanadium oxide layers, the amine spacer layers are embedded, as sketched in **c**. **d**, A typical powder X-ray diffraction pattern of the VO_x nanotubes, see Methods. **f**, The pattern is closely related to the $P4_2/m$ space group double-layered vanadium–oxygen structure of BaV₇O₁₆·*n*H₂O synthesized hydrothermally²⁷, in which vanadium atoms are in quasi-pyramidal (V(1), V(2)), and tetrahedral (V(3)) sites. **e**, A plan view of the top layer V(1) and V(2) edge-sharing square pyramids. The pyramids form zig-zagging chains linked via the corners of the V(1)-containing squares, while the tetrahedral V(3), nestled in the unoccupied ‘windows’ between the chains, stitches the two layers together (**e** and **f**). **g**, Energy hierarchy for the three vanadium sites obtained from the electron count calculated from the bond-valence sums.

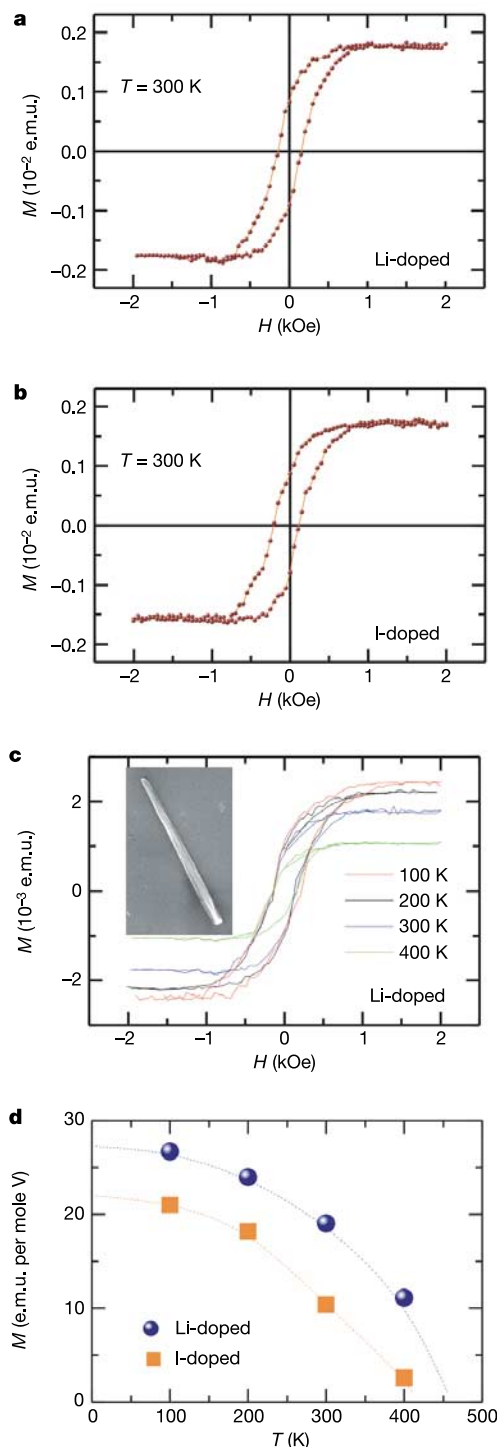


Figure 2 Magnetization of VO_x nanotubes after hole and electron doping. The tubular morphology appears intact after doping, as illustrated in the SEM image of a single ~85-nm-diameter lithiated tube (see ref. 28 for lithium intercalation into layered vanadates). Typical ferromagnetic magnetization M versus applied magnetic field H hysteresis (M - H) loops taken at room temperature after doping with lithium (electrons) (**a**) and doping with iodine (holes) (**b**). M - H loops are temperature-dependent (**c**), with the saturation moment characteristically decreasing with temperature, to be extinguished at the transition temperature $400 \text{ K} < T_c < 500 \text{ K}$ for either doping, as shown in **d**. The width of the loop of ~300 Oe is consistent with a dipolar origin²⁹ of the coercive field.

being the reflection in the mid-plane of the top layer, rotated by 90° ; an idealized diagram, not showing oxygen-bond distortions²⁷, is shown in Fig. 1e. Each layer consists of two octahedrally coordinated vanadium atoms, V(1) and V(2), with one tetrahedrally coordinated vanadium, V(3) lying in the mid-plane. The V(1) and V(2) sites have one short and one long oxygen bond, making the oxygen environment of these nominally octahedral sites closer to square pyramidal (Fig. 1f), with the pyramid apices pointing outward from the bilayer centre.

Electronically, the band structure of a similar system²² shows a narrow band associated with the lowest singlet crystal field level. Correspondingly, the electrostatic repulsion of the close-in pyramid apical oxygens at the V(1) and V(2) sites will favour a d_{xy} ground state, whereas in the tetrahedral V(3) site a $d_{x^2-y^2}$ ground state is expected, owing to the tetrahedral distortion (elongation along the z axis). We establish the energy hierarchy in the vanadium sites from the experimental parameters for the V–O bonds; the $\text{BaV}_7\text{O}_{16}$ -based²⁷ bond valence sum estimates yield fractional electron charges of 0.58, 0.82, and $1.06e^-$ on the V(1), V(2), and V(3) sites respectively. The high e^- occupations of V(2) and V(3) imply that these sites have the lowest energy (Fig. 1g), with the electrons relatively localized, whereas the high-lying V(1) sites are likely to be involved in itineracy and in changing occupation with doping. The deepest level, V(3), is also the most deeply buried in the structure, and the highest-lying V(1) level is also that of the most exposed site¹⁷, further supporting the idea that V(1) is the doping-sensitive site.

The effect of charge-doping the VO_x nanotubes is surprising. Figure 2a shows the lithiated case (see Methods), where the magnetization versus applied magnetic field (M – H) response is

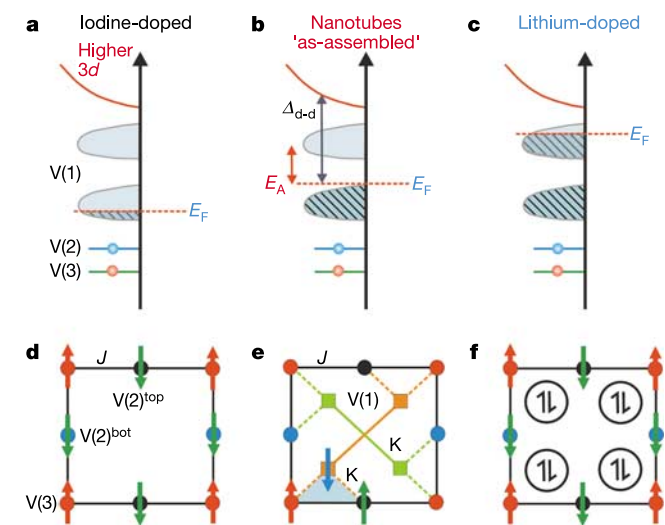


Figure 3 A schematic representation of Mott–Hubbard band splitting in VO_x nanotubes and a simple unit-cell model of spin textures with and without charge doping. In a Mott insulator, the motion of each particle is hindered by the strong Coulomb repulsion U from other particles, and in this respect they are very different from band insulators, in which the Pauli exclusion principle forbids the motion of electrons. **a–c**, V(2) and V(3) spins are in $3d_{xy}$ and $3d_{x^2-y^2}$ orbitals respectively, and split $3d_{xy}$ Hubbard bands of V(1), showing the Fermi-level E_F varying with doping (each component of the split bands contains $1e^-$ per site). If the Mott–Hubbard gap (controlled by U) is the transport gap (E_A , see text), it is optically forbidden, and the $\Delta_{d-d} \approx 1$ eV optical transition (see Supplementary Fig. S1, and ref. 30) is to the next highest (allowed on a site without symmetry elements), assumed itinerant component in the $3d$ band complex. In the unit-cell representation, lithium (**d**) and iodine (**f**) doping are magnetically equivalent. **e**, In the unit cell of as-assembled VO_x nanotubes two exchange interactions, K and J , acting in disharmony, are responsible (1) for spin frustration at the unit-cell edges (emphasized by a blue triangle), and (2) for maintaining the (zig-zagging) spin-dimers on top (orange squares) and bottom (green squares) layers of the tube wall.

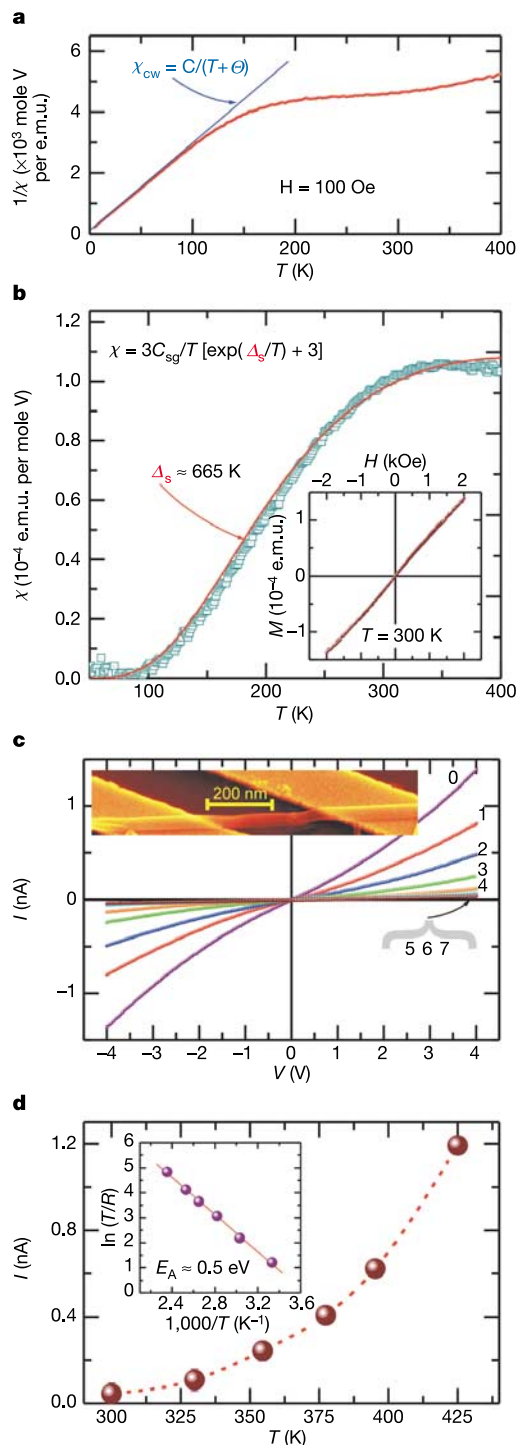


Figure 4 Spin-gapped magnetic susceptibility of as-assembled VO_x nanotubes, containing approximately one spin in each V(1), V(2), and V(3) site. **a**, Inverse susceptibility is linear in temperature at low T according to Curie–Weiss law (compatible with the V(3) spins). **b**, Spin susceptibility after subtracting a constant Van Vleck term and the χ_{CW} (see text) shows it has the behaviour characteristic of a spin-gap system, originating in the frustration introduced by the V(1) spins, as indicated in Fig. 3e. The inset in **b** shows that, in contrast to the Li- or I-doped tubes, the M – H loops (shown here at room temperature) are linear. **c**, Current–voltage (I – V) characteristics of individual VO_x tubes (inset displays the SEM image of the tube measured here) are nonlinear (hole-semiconducting), with the conductivity decreasing with Li doping (see Methods) increased (here in fixed steps from 0 to 7; where 0 corresponds to the as-assembled state). The (hopping) conduction is thermally activated (**d**), with the activation energy $E_A \approx 0.5 \text{ eV}$ (inset).

found to be nonlinear and hysteretic, saturating in the vicinity of ~ 1 kOe. The plot of saturation magnetization versus temperature (Fig. 2d) is typical of a ferromagnet with a Curie (transition) temperature of $T_C \approx 500$ K—well beyond room temperature—showing that these materials pass the test of being magnetic at room temperature (essential for practicable device applications). Even more remarkable, removing electrons (by hole-doping using iodine) results in a nearly identical spin response (Fig. 2b, d)—indicating an unexpected electron–hole complementarity that we have discovered in these nanotube structures.

This charge complementarity suggests that the key to understanding the spin response of the VO_x nanotubes lies in the modification in electronic occupation of the V(1) level as a function of doping. In the as-assembled VO_x nanotubes, owing to protonation of the amine NH_3^+ head groups²⁶, there is a stoichiometry-based electron doping of $0.87e^-$ per V(1) site. The addition of Li increases this to 1.85 per V(1). Saturation with iodine will reduce the electron count significantly, and the charge transfer will have opposite sign but a similar magnitude to that in the lithiated case, reducing the V(1) occupation to a low value.

In a view consistent with the magnetic data, in the as-assembled state there is approximately one electron in all the lowest crystal field levels, whereas the doped systems have either zero (iodine) or two electrons (lithium) in the V(1), retaining one electron in the other levels (V(2) and V(3)). In this view, iodine doping (introducing holes) is the simplest to visualize: only the V(2) and V(3) levels are occupied; if the occupations are indeed nearly integral these occupied 3d V(2) and V(3) levels should behave as spins. These spins will interact via a nearest-neighbour exchange J (see Fig. 3). If the V(1) is empty, then the top-layer V(2) sites form spin chains (...red-yellow-red-yellow...; see Fig. 1e) with the mid-plane V(3) sites, linked to an orthogonal set of V(2)–V(3) chains in the bottom layer, forming a two-dimensional (2D) magnetic network. When J is antiferromagnetic, as is usual in oxides^{14,22,23}, then the V(2) spins will have opposite polarity to the V(3) spins (ferrimagnetism), giving one spin per seven vanadiums. (For ferromagnetic J , the system would be a ferromagnet with three spins per formula unit.) In the lithiated case, the hysteresis loop is remarkably similar to the hole-doped one, which is explained by the symmetry between doubly occupied and empty V(1) sites, of which neither contributes to the moment. The observed saturation moment (Fig. 2b) is $\sim 1/30$ of that expected in the strong coupling ferrimagnetic picture outlined above. Although weak (possibly itinerant) ferromagnetism is observed in some vanadates²⁴, it is more likely that in the nanotubes, inhomogeneities in the tube dimensions (see Fig. 1) and in the dopant penetration into the inter-wall spaces will reduce the measured moment.

The physics underlying the spin behaviour of undoped nanotubes is very different. Figure 4a shows the measured static linear magnetic susceptibility $\chi = M/H$ of a randomly oriented collection of tubes. At temperatures beyond ~ 100 K, the main features of the χ – T curves clearly do not correspond to a simple Curie–Weiss law, reflecting the low-dimensionality of the VO_x tubes. χ can be written as a sum of three parts. One is the temperature-independent (diamagnetic plus Van Vleck paramagnetic orbital) part, $\chi_0 = 2.62 \times 10^{-5}$ e.m.u. per mole V, where e.m.u. are electromagnetic units, a value in good correspondence with other layered vanadates^{19,25}. Another is a contribution from spins obeying the low-temperature Curie–Weiss law, $\chi_{\text{CW}}(T) = C/(T + \Theta)$ (here $C \approx 3.58 \times 10^{-2}$ e.m.u. per mole V is a Curie constant, and the nearly null Weiss parameter Θ is an indicator of null long-range magnetic order). We estimate the effective moment $p_{\text{eff}} = (3k_B C/N\mu_B^2)^{1/2}$ of spins contributing to the low- T χ_{CW} at about $0.5 \mu_B$ (N is the particle density, k_B is the Boltzmann constant, and μ_B is the Bohr magneton) and argue below that this is consistent with the contribution from spins on the V(3) sites. After subtracting χ_0 and χ_{CW} , what is left is χ_d , plotted in Fig. 4b. Displaying a broad

maximum (somewhat beyond room temperature), χ_d is a ‘thumb-print’ of a low-dimensional quantum spin liquid¹⁵ with a gap in the spin-excitation spectrum. Indeed, the simplest picture of dimerized spins thermally excited over a spin singlet–triplet gap, $\chi_d = N' g^2 \mu_B^2 / k_B T [\exp(\Delta_s/k_B T) + 3]$ gives a good fit to the data with the gap $\Delta_s \approx 665$ K = 57 meV, comparable to the gap found in CaV_2O_5 , for example; see ref. 25. Here N' and $g \approx 1.97$ denote the number and the Landé g -factor of the contributing V^{4+} ions.

The effect of the V(1) spins in the undoped system is illustrated in Fig. 3e, where K is a nearest-neighbour exchange coupling between the V(1) spin and its nearest neighbours. With the V(2)–V(3) interaction J expectedly antiferromagnetic, K acts to frustrate the antiferromagnetic spin alignment between V(3) and V(2), thus destroying the (ferri)magnetic order. The chain-like V(1)–V(2) spin structures (Fig. 1e) in the upper and lower planes are compatible with a spin-gap (or spin-liquid) state^{22,25}; the V(3) spins could be isolated from the biplanar spin-liquid wavefunctions, resulting in the observed large Curie term (leaving the spin liquid with an even number of electrons per cell, as in the $S = 1/2$ Heisenberg chain¹⁵, for example).

The schematic energy-band diagrams in Fig. 3a–c illustrate band splitting due to the Coulomb repulsion in the V 3d orbitals^{14,18}. The half-filled Mott insulating band describes a spin that is robust against either adding or removing charge, owing to the Mott gap. In the as-assembled nanotubes, we interpret all bands as having approximately unit occupation, making all V sites spin-like; in the lithiated system the V(1) bands are full, or in the iodated state empty, making the V(1) sites spinless in either case, which removes the frustration of the V(2)–V(3) interactions J and permits ferrimagnetism. The T_C of the ferrimagnetic state and the spin gap energy gap in the undoped state are comparable, indicating their common origin in relatively large exchange interactions. Transport measurements on the individual tubes show nonlinear current–voltage characteristics, with the conductivity thermally activated (Fig. 4c, d) over the barrier $E_A \approx 0.5 \pm 0.1$ eV. The charge carriers participating in the conduction process appear to be holes, not unlike the carriers in the layered copper oxides¹⁵. The apparent p -type character of the as-assembled tubular VO_x manifests as a relative decrease of the conductance upon doping with lithium, as shown in Fig. 4c. The hole-like transport in the undoped nanotubes places the Fermi level in the lower Hubbard band (Fig. 3b), as is compatible with a V(1) occupation of less than $1e^-$.

Thus, the emerging picture is of open-net layered-wall vanadium oxide nanotubes that can easily be spin-tuned by charge doping. In doping, the Fermi level is swept through the Mott gap, removing the frustration in the undoped system responsible for the spin gap, and promoting ferrimagnetism. Itinerant carriers under spin control are produced in the V(1) band, which interacts with the other more localized vanadium spins. Although manipulating and achieving doping selectivity of these tubes deserve further effort towards the practical implementation of this effect, our findings suggest a path to new spin-aligned nanoscale building blocks, in which the Fermi-level sweep can be accomplished by applied voltage. $\square$

Methods

VO_x nanotubes were synthesised hydrothermally¹¹ using vanadium(V) triisopropoxide, $\text{VO}(\text{iOPr})_3$, and a dodecylamine, solvated in a 2:1 molar ratio in absolute ethanol (3 ml per g of $\text{VO}(\text{iOPr})_3$), stirred at first under nitrogen atmosphere for roughly 1 h. The resulting light-yellow solution of alcoxide-amine adduct was hydrolysed with H_2O (5 ml per g of $\text{VO}(\text{iOPr})_3$) while stirring. The adduct was continually stirred and aged at room temperature for ~ 4 days, turning dark orange. A reaction at 180°C in a Teflon-lined Parr acid digestion vessel for ~ 8 days yielded a ‘black’ product, which was then washed with ethanol and hexane and dried at 85°C in vacuum for ~ 30 h. Scanning (SEM) and transmission (TEM) electron microscopies showed that the end-product were nanotubes (see Fig. 1). X-ray diffraction (Siemens; $\text{Cu K}\alpha$ radiation, $\lambda = 1.5418 \text{ \AA}$) displayed three relatively strong peaks at 27° [001], 13.4° [002], and 9.0° [003], in accord with refs 11 and 26. This pattern is indexed with a tetragonal basal plane with $a = 6.144 \text{ \AA}$ and $c = 26.6 \text{ \AA}$. This size a parameter and the square basal plane lattice comes from a double sheet of vanadium atoms, containing vanadium mainly in a distorted octahedral

configuration. Lithium doping was done with 1.69 mol BuLi solution (0.34 ml per 0.25 g of VO_x nanotubes, calculated to compensate each V^{5+} with one Li). Iodine doping was done by submitting VO_x nanotubes to iodine vapour for varying time intervals (0.5, 1, 1.5 and 2 h) from iodine crystals (0.125 g of VO_x per 0.85 g of I) in evacuated (1.3×10^{-6} torr) and sealed 125-cm³ quartz tubes. After 1 h, the nanotubes visibly darkened, with no measurable change afterwards. All chemicals were from Aldrich. For magnetic measurements, doped nanotubes were weighed and sealed under nitrogen in gel capsules to avoid dopant oxidation (see ref. 8). Magnetic measurements were performed in a SQUID magnetometer (MPMS, Quantum Design) in fields up to 5 T. For two-probe transport measurements, the tubes were cast out of hexane solution onto oxidized silicon substrates, which were e-beam prepatterned with Pt/Ti or Au/Ti electrodes. The oxide thickness was 140 nm and the electrodes were 300 nm wide, with separation varied in the 200–800 nm range. All transport measurements were performed in vacuum (starting at 3.8×10^{-7} torr and prebaking for 17 h to improve contact resistance), with the *in situ* Li doping from a commercial (SAES Getter) source, following an alkali-doping process used for carbon nanotubes⁸. Li was added in fixed steps; in each step the source was heated by applying current $I = 6.7$ A for 1.5 min (see Fig. 4c).

Received 19 May; accepted 25 August 2004; doi:10.1038/nature02970.

- Iijima, S. Helical microtubules of graphitic carbon. *Nature* **354**, 56–58 (1991).
- Cui, Y. & Lieber, C. M. Functional nanoscale electronic devices assembled using silicon nanowire building blocks. *Science* **291**, 851–853 (2001).
- Tenne, R., Margulis, L., Genut, M. & Hodes, G. Polyhedral and cylindrical structures of tungsten disulphide. *Nature* **360**, 444–446 (1992).
- Rosenfeld-Hacohen, Y., Grunbaum, E., Tenne, R., Sloan, J. & Hutchison, J. L. Cage structures and nanotubes of NiCl_2 . *Nature* **395**, 336–337 (1998).
- Remskar, M. et al. Self-assembly of subnanometer-diameter single-wall MoS_2 nanotubes. *Science* **292**, 479–481 (2001).
- Dekker, C. Carbon nanotubes as molecular quantum wires. *Phys. Today* **52** (5), 22–28 (1999).
- Fuhrer, M. S. et al. Crossed nanotube junctions. *Science* **288**, 494–497 (2000).
- Derycke, V., Martel, R., Appenzeller, J. & Avouris, Ph. Carbon nanotube inter- and intramolecular logic gates. *Nano Lett.* **1**, 453–456 (2001).
- Wolf, S. A. et al. Spintronics: A spin-based electronics vision for the future. *Science* **294**, 1488–1495 (2001).
- Hueso, L. & Mathur, N. Dreams of a hollow future. *Nature* **427**, 301–303 (2004).
- Krumeich, F. et al. Morphology and topochemical reactions of novel vanadium oxide nanotubes. *J. Am. Chem. Soc.* **121**, 8324–8331 (1999).
- Tsukagoshi, K., Alphenaar, B. W. & Ago, H. Coherent transport of electron spin in a ferromagnetically contacted carbon nanotube. *Nature* **401**, 572–574 (1999).
- Yao, Z., Postma, H. W. Ch., Balents, L. & Dekker, C. Carbon nanotube intramolecular junctions. *Nature* **402**, 273–276 (1999).
- Tokura, Y. & Nagaosa, N. Orbital physics in transition-metal oxides. *Science* **288**, 462–468 (2000).
- Imada, M., Fujimori, A. & Tokura, Y. Metal-insulator transitions. *Rev. Mod. Phys.* **70**, 1039–1263 (1998).
- Levy, P., Leyva, A. G., Troiani, H. E. & Sánchez, R. D. Nanotubes of rare-earth manganese oxides. *Appl. Phys. Lett.* **83**, 5247–5249 (2003).
- Zavalij, P. Y. & Whittingham, M. S. Structural chemistry of vanadium oxides with open frameworks. *Acta Cryst. B* **55**, 627–663 (1999).
- Mott, N. F. *Metal-Insulator Transitions* (Taylor & Francis, London, 1974).
- Kanada, M. et al. On the magnetic properties of systems with low dimensional linkage of VO_5 pyramids. *J. Phys. Soc. Jpn* **67**, 2904–2909 (1998).
- Limelette, P. et al. Universality and critical behavior at the Mott transition. *Science* **302**, 89–92 (2003).
- Yamauchi, T., Ueda, Y. & Mori, N. Pressure-induced superconductivity in $\beta\text{-Na}_{0.33}\text{V}_2\text{O}_5$ beyond charge ordering. *Phys. Rev. Lett.* **89**, 057002 (2002).
- Pickett, W. E. Impact of structure on magnetic coupling in CaV_4O_9 . *Phys. Rev. Lett.* **92**, 056402 (2004).
- Korotin, M. A. et al. Exchange interactions and magnetic properties of the layered vanadates CaV_2O_5 , MgV_2O_5 , CaV_3O_7 , and CaV_4O_9 . *Phys. Rev. Lett.* **83**, 1387–1390 (1999).
- Lumsden, M. D., Sales, B. C., Mandrus, D., Nagler, S. E. & Thompson, J. R. Weak ferromagnetism and field-induced spin reorientation in $\text{K}_2\text{V}_3\text{O}_8$. *Phys. Rev. Lett.* **86**, 159–162 (2001).
- Onoda, M. & Nishiguchi, N. Crystal structure and spin gap state of CaV_2O_5 . *J. Solid-State Chem.* **127**, 359–362 (1996).
- Dobley, A. et al. Manganese vanadium oxide nanotubes: synthesis, characterization, and electrochemistry. *Chem. Mater.* **13**, 4382–4386 (2001).
- Wang, X., Liu, L., Bontchev, R. & Jacobson, A. J. Electrochemical-hydrothermal synthesis and structure determination of a novel layered mixed-valence oxide: $\text{BaV}_7\text{O}_{16}\cdot\text{nH}_2\text{O}$. *J. Chem. Soc. Chem. Commun.* 1009–1010 (1998).
- Bergström, Ö., Gustasson, T. & Thomas, J. O. Electrochemically lithiated vanadium oxide, $\text{Li}_x\text{V}_6\text{O}_{13}$. *Acta Cryst. C* **53**, 528–530 (1997).
- Bertotti, G. *Hysteresis in Magnetism* (Academic, London, 1998).
- Cao, J. et al. Effect of sheet distance on the optical properties of vanadate nanotubes. *Chem. Mater.* **16**, 731–736 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Afzali, K.-S. Cho, C. R. Kagan, F. X. Redl and S. Sun for technical advice, C. A. Feild for chemistry insights, P. Y. Zavalij for his expertise in crystal structures, B. Spivak and A. M. Tsvelik for discussions, and R. Ludeke for his contributions. This work is supported in part by the Defense Advanced Research Project Agency (DARPA).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.K.-E. (krusin@us.ibm.com).

Biogenically driven organic contribution to marine aerosol

Colin D. O'Dowd^{1*}, Maria Cristina Facchini^{2*}, Fabrizia Cavalli², Darius Ceburnis^{1,3}, Mihaela Mircea², Stefano Decesari², Sandro Fuzzi², Young Jun Yoon¹ & Jean-Philippe Putaud⁴

¹Department of Experimental Physics and Environmental Change Institute, National University of Ireland, Galway, Ireland

²Istituto di Scienze dell'Atmosfera e del Clima – Consiglio Nazionale delle Ricerche, 40129 Bologna, Italy

³Environmental Physics and Chemistry Division, Institute of Physics, Savanoriu 231, LT - 2053 Vilnius, Lithuania

⁴European Commission, Institute for Environment and Sustainability, Joint Research Centre, 21021 Ispra, Italy

* These authors contributed equally to this work

Marine aerosol contributes significantly to the global aerosol load and consequently has an important impact on both the Earth's albedo and climate. So far, much of the focus on marine aerosol has centred on the production of aerosol from sea-salt¹ and non-sea-salt sulphates^{2,3}. Recent field experiments, however, have shown that known aerosol production processes for inorganic species cannot account for the entire aerosol mass that

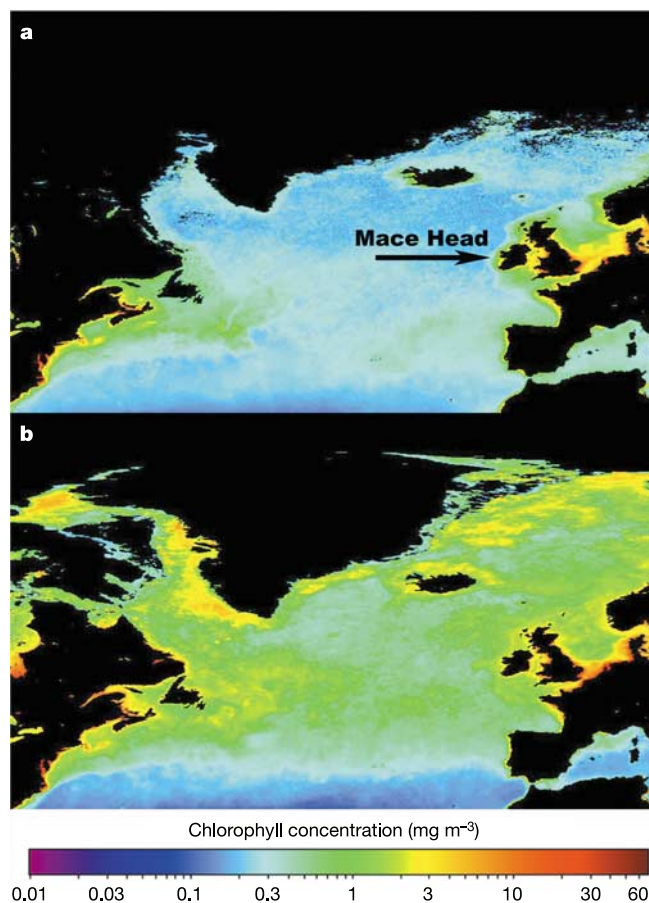


Figure 1 Organic matter at the sea surface. SeaWiFS-derived seasonal average (5-year) sea-surface chlorophyll concentrations in winter (a) and spring (b), illustrating low biological activity in North Atlantic waters during winter and high activity in spring (courtesy of SeaWiFS Project, NASA/Goddard Space Flight Center and ORBIMAGE). The location of Mace Head is shown in a. Marine air masses arrive at Mace Head after at least 96 h transit over the ocean from the Arctic and northwest Atlantic.

occurs in submicrometre sizes^{4–6}. Several experimental studies have pointed to the presence of significant concentrations of organic matter in marine aerosol^{7–11}. There is some information available about the composition of organic matter^{12–14}, but the contribution of organic matter to marine aerosol, as a function of aerosol size, as well as its characterization as hydrophilic or hydrophobic, has been lacking. Here we measure the physical and chemical characteristics of submicrometre marine aerosol over the North Atlantic Ocean during plankton blooms progressing from spring through to autumn. We find that during bloom periods, the organic fraction dominates and contributes 63% to the submicrometre aerosol mass (about 45% is water-insoluble and about 18% water-soluble). In winter, when biological activity is at its lowest, the organic fraction decreases to 15%. Our model simulations indicate that organic matter can enhance the cloud droplet concentration by 15% to more than 100% and is therefore an important component of the aerosol–cloud–climate feedback system involving marine biota.

During the year 2002 we performed measurements of size-resolved physical and chemical properties of aerosols collected in northeast Atlantic marine air arriving at the Mace Head Atmospheric Research Station, on the west coast of Ireland. Air masses arriving at Mace Head under clean-sector selection criteria¹⁵ typically spent four days advecting over the North Atlantic Ocean, more often emerging from the Canadian, Greenland and Arctic regions and less frequently from northern USA and subtropical

regions¹⁵. Analysis of available data indicates that coastal influences are negligible (see Supplementary Information).

Both the aerosol physical and chemical properties exhibited clear seasonal patterns following biological activity in the North Atlantic. Analysis of SeaWiFS-derived seawater chlorophyll, an indicator of biological activity, shows high chlorophyll concentrations from spring through to autumn, corresponding to the North Atlantic plankton blooming period, and minimum concentrations occurring in winter. Figure 1 illustrates a roughly tenfold increase in chlorophyll from winter to spring. Over the year, the period of low biological activity (LBA) can be regarded as winter and the period of high biological activity (HBA) can be regarded as spring through to autumn.

The aerosol chemical data were split into these two periods for analysis and a marked contrast between them was found with respect to the relative concentrations of inorganic salts, total organic carbon (TOC), water-soluble organic carbon (WSOC), and water-insoluble organic carbon (WIOC), particularly for the Aitken and accumulation size ranges. The average relative chemical compositions and mass distributions for the LBA and HBA periods are illustrated in Fig. 2, where the concentration of carbon classes are reported as organic mass. The size fractions from 0.125 to 0.5 μm cover the accumulation mode and the size fraction from 0.06 to 0.125 μm covers most of the Aitken mode.

During the LBA period, sea-salt aerosol dominated all size fractions with a 74% ($0.3 \mu\text{g m}^{-3}$) contribution to the accumulation

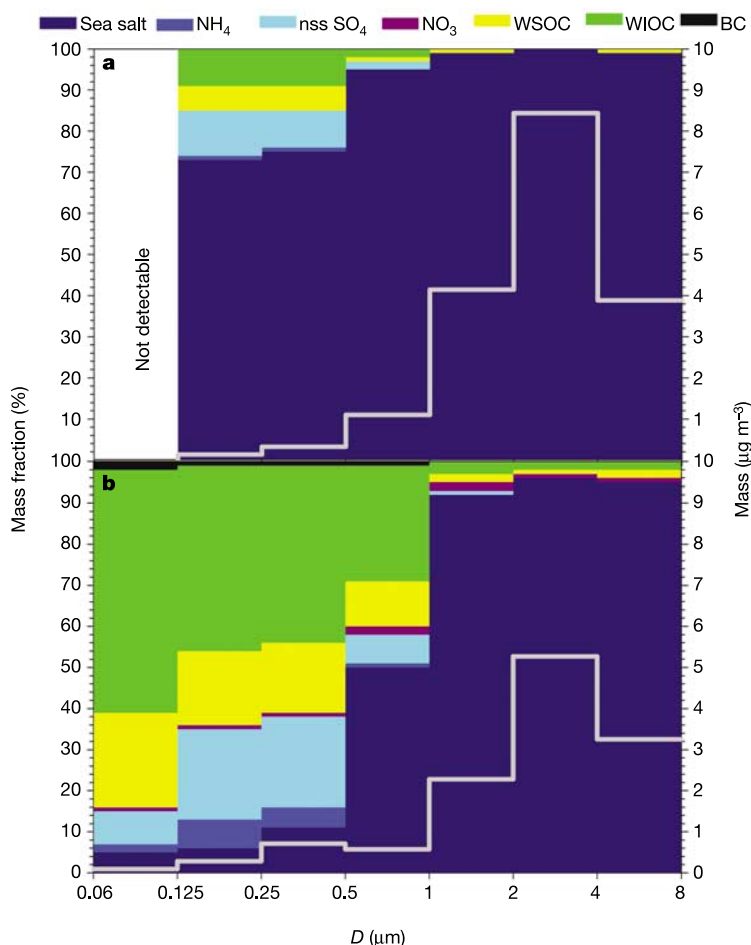


Figure 2 Chemical composition of marine aerosols. Shown are average size-segregated chemical compositions and absolute mass concentrations for North Atlantic marine aerosols sampled with a Berner Impactor, for LBA (a) and HBA (b) periods. The

concentrations of WSOC, WIOC and BC are reported as mass of organic matter (see ref. 15 for a full discussion).

mode mass. The remainder of the mass in this mode comprised 10% ($0.045 \mu\text{g m}^{-3}$) non-sea-salt (nss) sulphate and 15% ($0.070 \mu\text{g m}^{-3}$) TOC. By contrast, in the HBA period, the TOC fraction increased markedly, particularly for the submicrometre sizes, and this fraction increased with decreasing size. For the HBA period, TOC in the accumulation mode contributed 65% ($0.619 \mu\text{g m}^{-3}$) of the mass, whereas sea salt contributed 10% ($0.097 \mu\text{g m}^{-3}$) and nss sulphate 23% ($0.216 \mu\text{g m}^{-3}$). The largest percentage contribution (83%) of TOC occurred in the fine mode ($0.06\text{--}0.125 \mu\text{m}$). Across the size range from 0.06 to $0.5 \mu\text{m}$, the average WIOC and WSOC contributions were 45% and 18%, respectively.

Detailed analysis of WSOC aerosol revealed the presence of partly oxidized species with extended aliphatic moieties, partly attributable to high-molecular-mass species. Such compounds are responsible for the large decrease in surface tension observed in aerosol water extracts¹⁵. Surface-active organic matter (OM) of biogenic origin (such as lipidic and proteinaceous material and humic substances), enriched in the oceanic surface layer and transferred to the atmosphere by bubble-bursting processes, are the most likely candidates to contribute to the observed organic fraction in marine aerosol^{16,17}. Such surface-layer enrichment of OM is readily seen in terms of chlorophyll concentrations illustrated in Fig. 1.

The observed organic aerosol characteristics are consistent with laboratory studies on aerosol generated from Atlantic sea water¹⁶ that showed a peak in organic aerosol concentration, and a concomitant increase in WIOC and high-molecular-mass surface-active fractions, during periods of blooming phytoplankton. Moreover, the increasing enrichment of the aerosol organic fraction with decreasing size is consistent with thermodynamic predictions¹⁸ of

bubble-bursting processes under conditions in which the ocean surface layer becomes concentrated with surfactant material that can be incorporated into sea-spray drops in addition to inorganic salts¹⁹.

Accumulation-mode and Aitken-mode size-distribution properties also reveal seasonal behaviour. Both modes were log-normally distributed, with a minimum mean size in winter and a maximum in summer. For the Aitken mode, the mean modal diameter increased from 28 nm in winter to 46 nm in summer (Fig. 3). Similarly, the mean modal diameter for the accumulation mode increased from 97 nm in winter to 164 nm (maximum 195 nm) in summer. As the modal diameter increased, the width of the distribution also narrowed. The average concentration for the accumulation mode remained constant in both periods ($N \approx 120 \text{ cm}^{-3}$), whereas the concentration for the Aitken mode increased from about 150 cm^{-3} in the LBA period to about 300 cm^{-3} in the HBA period. The increase in mean aerosol size is primarily driven by the increase in TOC and could be attributable to either primary or secondary production processes. Secondary aerosol formation should be demonstrated by modal growth over timescales of $12\text{--}24 \text{ h}$, particularly under conditions of high solar radiation; however, such growth was not observed in the cases studied. If the aerosol OM arose from condensation, then a twofold or threefold more OM should condense on super-micrometre sizes¹ because of the dominant super-micrometre condensation sink; however, the opposite was observed. Further, it is unlikely that the dominant WIOC could be formed through cloud processing.

Primary production, in contrast, can also lead to such increases in mean size: during the HBA period, the surface layer is enriched in OM that becomes adsorbed on bursting bubbles. During the bubble-bursting process that produces spray droplets, the OM can displace water in the spray droplet and lead to a larger residual aerosol size when the droplet water evaporates. At typical North Atlantic wind speeds of 10 m s^{-1} , the primary marine aerosol flux²⁰ is about $2 \times 10^6 \text{ m}^{-2} \text{ s}^{-1}$. Thus, after transit over the Atlantic, primary aerosol generation can contribute an input of about

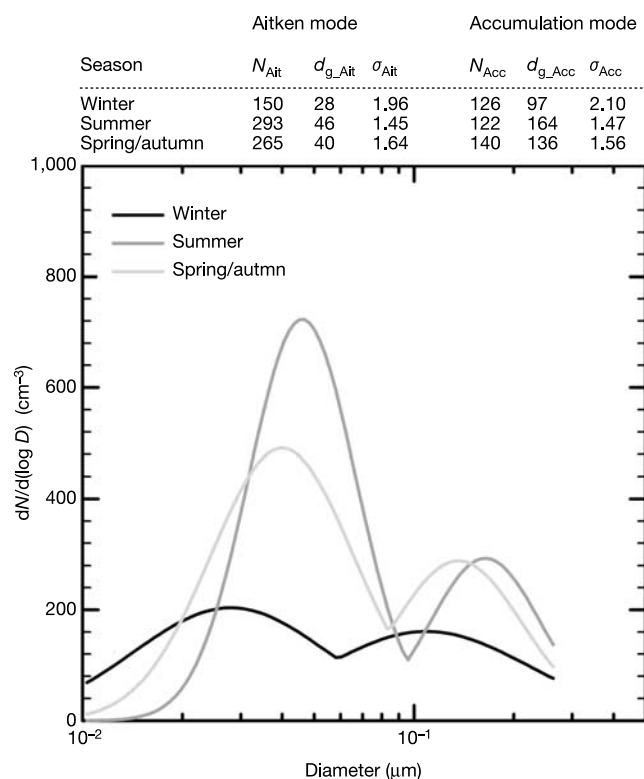


Figure 3 Seasonal characteristics of aerosol microphysics. North Atlantic Aitken-mode and accumulation-mode log-normal aerosol size distributions are shown, based on average fitted log-normal distributions for winter, summer and spring/autumn. For clarity, the log-normal modes are truncated in regions of overlap between the Aitken and accumulation modes. Log-normal parameters are given for each mode (N is the mode concentration, d_g is the mean mode size, and σ is the standard deviation of the log-normal distribution).

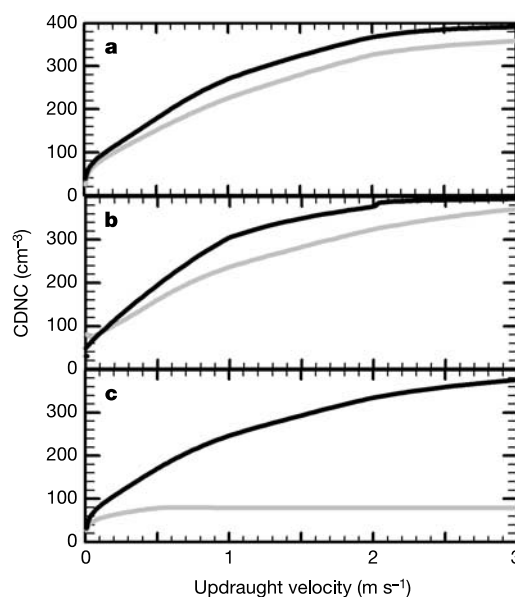


Figure 4 Increase in CDNC due to the addition of OM. **a**, CDNC as a function of internally mixed sulphate and sea salt (base case) and as a function of OM internally mixed with sulphate and sea salt (base case plus OM). **b**, CDNC as a function of an external mixture of sulphate and sea salt (base case) and of OM internally mixed with sea salt (base case plus OM). **c**, Internal mixture of sulphate and sea salt (base case) and OM externally mixed (base case plus OM). In all panels the black line shows the base case plus OM, and the grey line the base case.

700 particles cm^{-3} into the marine boundary layer aerosol. Notwithstanding removal processes for both continental and marine aerosol, a primary source can easily explain the observed number concentrations and organic fraction in clean marine air.

Thus, the combined aerosol physicochemical characteristics indicate a significant primary aerosol source in which bubble-bursting produces submicrometre particles²¹ enriched in insoluble and high-molecular-mass OM^{16,19} and in which the increase in particle size derives from an increased absorption of surface-active organics onto bubbles as the concentration of biogenic OM in the surface layer increases with biological activity. It should be noted that secondary aerosol formation involving nss sulphate also increases during the HBA period, and it is seen as an increase in both total number concentration and nss sulphate mass during this period; however, the increase is noticeably less than that of OM.

Because the population of cloud condensation nuclei (CCN) is typically dominated by submicrometre-sized particles, the data presented here completely change the picture of what influences marine CCN given that WSOC, WIOC and surface-active properties, all of which influence the CCN activation potential^{22,23}, are typically not parameterized in current climate models. To evaluate the impact of this organic aerosol source on cloud microphysics, cloud droplet number concentration (CDNC) has been calculated with a cloud model (see Methods) for inorganic-only CCN (base case) and inorganic plus organic CCN (base case plus organics). The mixing state of the aerosol population can influence the predicted CDNC; however, because this is not known, simulation results are presented for both internally and externally mixed populations.

CDNCs for three state-of-mixing models are presented (see Methods): first, one in which all species are internally mixed; second, one in which sulphate is externally mixed whereas sea salt and OM are internally mixed; and last, one in which sulphate and sea salt are internally mixed and OM is treated as externally mixed. The input size distributions are constrained by the combined size and chemical distributions observed in the HBA period. For both internally mixed models, the addition of WSOC mass, despite the large WIOC fraction, increases CDNC by about 15–20% (Fig. 4) over an updraught range of 0.1–3 m s^{-1} . By comparison, when OM is treated as externally mixed aerosol, CDNC increases from about 29% at 1 m s^{-1} to more than 100% at 3 m s^{-1} , thus showing a significant impact of OM on CDNC. These results have significant implications for the prediction of cloud albedo and cloud lifetime, both of which are influenced by the cloud droplet population, which in turn is driven by CCN properties. Further work is required to determine the impact of complex mixed organic–inorganic CCN on cloud properties.

Our results indicate that an important source of OM from the ocean is omitted from current climate-modelling predictions and should be taken into account. Furthermore, the production of WSOC and WIOC in surface waters, their distribution in the surface layer and their ultimate transfer to the aerosol phase warrants active research to provide a full explanation of their relative contributions to the marine aerosol as a function of biological activity. Given that the evolution of micro-algae is driven by environmental change, particularly increasing oceanic temperatures²⁴, the production of primary organic marine aerosol represents a newly identified and potentially important component of the marine biota and climate feedback system involving aerosols and clouds^{2,3}. □

Methods

Aerosol measurements

Size distributions (10–300 nm) were taken with a scanning mobility particle sizer, which derived size distributions every 2 min. Clean air periods were screened from the continuous measurements at Mace Head during 2002. A total of 25 periods over the year were analysed for size distribution properties. The periods included in the data set covered 12–36-h sampling in clean air. The average size distribution of each period was fitted for Aitken-mode and accumulation-mode log-normal parameters; each period was then categorized by season. Nucleation modes, resulting from coastal particle formation events,

were excluded from the analysis. The chemical composition measurements covered typically 7 days' intermittent sampling in the cleanest northeast Atlantic air. Seven such samples comprised the HBA period average, and six such samples comprised the LBA period average. A full description of sampling procedures and criteria is presented in ref. 15 as well as an evaluation of the possible contribution of sampling artefacts to the data presented. It was found that if sampling artefacts were present, they were typically less than the level of random uncertainty associated with the measurements.

Cloud droplet model

The cloud model is a one-dimensional lagrangian cloud-parcel model with explicit microphysics, activation, and growth of a droplet population as a function of vertical velocity, temperature, pressure, water vapour mixing ratio, liquid water mixing ratio and saturation ratio inside the rising air parcel. The equilibrium saturation is computed with a modified form of the Köhler equation and the cloud droplet surface tension variation as a function of dissolved organic carbon concentration. The model is based on that reported in ref. 23, and references therein, but with the addition of segregated chemistry over internally and externally mixed aerosols. The aerosol size distribution function is modelled for the air parcel by using a sectional representation. Three hundred size sections, equally spaced on a logarithmic scale, are created between 0.01 and 16 μm . The growth of the aerosol particles due to water vapour uptake is calculated on a moving grid to eliminate numerical diffusion. This means that during the simulations the aerosol particles keep their initial bin number, but their wet diameter varies with water condensation or evaporation. In all simulations performed, the air parcel started rising at 98% relative humidity, 273 K and 800 mbar.

Aerosol mixing state input to simulations

Model 1 (sulphate–sea-salt–OM internal mix). The base case comprises an internal mix of sulphate and sea salt, whereas the base case plus OM comprises an internal mix of sulphate, sea salt and OM. In all models treated, the Aitken mode and the accumulation mode have the following fixed log-normal σ parameters: $\sigma_{\text{Aitken}} = 1.35$ and $\sigma_{\text{accumulation}} = 1.49$. For the base case, $d_{\text{Aitken}} = 30 \text{ nm}$ and $d_{\text{accumulation}} = 100 \text{ nm}$; $N_{\text{Aitken}} = 300 \text{ cm}^{-3}$ and $N_{\text{accumulation}} = 100 \text{ cm}^{-3}$. When OM is added as an internal mix (base case plus OM), N and σ remain constant but d_{Aitken} increases to 70 nm and $d_{\text{accumulation}}$ increases to 200 nm.

Model 2 (sulphate externally mixed; sea salt and OM internally mixed). The base case comprises separate sulphate and sea-salt modes. For sulphate, $d_{\text{Aitken}} = 70 \text{ nm}$ and $d_{\text{accumulation}} = 200 \text{ nm}$; $N_{\text{Aitken}} = 20 \text{ cm}^{-3}$ and $N_{\text{accumulation}} = 20 \text{ cm}^{-3}$. For sea salt, $d_{\text{Aitken}} = 30 \text{ nm}$ and $d_{\text{accumulation}} = 100 \text{ nm}$; $N_{\text{Aitken}} = 280 \text{ cm}^{-3}$ and $N_{\text{accumulation}} = 80 \text{ cm}^{-3}$. When OM is added as an internal mix with sea salt (base case plus organics), only the sea-salt modal diameter changes, resulting in $d_{\text{Aitken}} = 70 \text{ nm}$ and $d_{\text{accumulation}} = 200 \text{ nm}$.

Model 3 (sulphate and sea salt internally mixed; OM externally mixed). The base case comprises an internal mix of sulphate and sea salt, and $d_{\text{Aitken}} = 70 \text{ nm}$ and $d_{\text{accumulation}} = 200 \text{ nm}$; $N_{\text{Aitken}} = 42 \text{ cm}^{-3}$ and $N_{\text{accumulation}} = 37 \text{ cm}^{-3}$. OM is then added to the aerosol population as an external mix with size distribution parameters $d_{\text{Aitken}} = 70 \text{ nm}$ and $d_{\text{accumulation}} = 200 \text{ nm}$; $N_{\text{Aitken}} = 258 \text{ cm}^{-3}$ and $N_{\text{accumulation}} = 63 \text{ cm}^{-3}$.

Received 5 May; accepted 6 August 2004; doi:10.1038/nature02959.

- O'Dowd, C. D., Lowe, J. A. & Smith, M. H. Marine aerosol, sea-salt, and the marine sulphur cycle: A short review. *Atmos. Environ.* **31**, 73–80 (1997).
- Shaw, G. Bio-controlled thermostat involving the sulfur cycle. *Clim. Change* **5**, 297–303 (1983).
- Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulfur, cloud albedo and climate. *Nature* **326**, 655–661 (1987).
- McInnes, L. M., Quinn, P. K., Covert, D. S. & Anderson, T. L. Gravimetric analysis, ionic composition and associated water mass of the marine aerosol. *Atmos. Environ.* **30**, 869–884 (1996).
- Hubert, B. J. *et al.* Filter and impactor measurements of anions and cations during the First Aerosol Characterization Experiment (ACE 1). *J. Geophys. Res.* **103**, 16493–16509 (1998).
- Quinn, P. K. *et al.* Surface submicron aerosol chemical composition: What fraction is not sulfate? *J. Geophys. Res.* **105**, 6785–6805 (2000).
- Hoffman, E. J. & Duce, R. A. The organic carbon content of marine aerosol collected on Bermuda. *J. Geophys. Res.* **79**, 4474–4477 (1976).
- Novakov, T. *et al.* Organic aerosols in the Caribbean trade winds: A natural source? *J. Geophys. Res.* **102**, 21307–21313 (1997).
- Middlebrook, A. M., Murphy, D. M. & Thomson, D. Observation of organic material in individual marine particles at Cape Grim during the First Aerosol Characterization Experiment (ACE 1). *J. Geophys. Res.* **103**, 16475–16483 (1998).
- Putaud, J.-P. *et al.* Chemical mass closure and assessment of the origin of the submicron aerosol in the marine boundary layer and the free troposphere at Tenerife during ACE-2. *Tellus* **52B**, 141–168 (2000).
- Kleefeld, S., Hoffer, A., Krivacsy, Z. & Jennings, S. G. Importance of organic and black carbon in atmospheric aerosols at Mace Head, on the West Coast of Ireland. *Atmos. Environ.* **36**, 4479–4490 (2002).
- Kawamura, K. & Gagosian, R. B. Mid-chain ketocarboxylic acids in the remote marine atmosphere: Distribution patterns and possible formation mechanisms. *J. Atmos. Chem.* **11**, 107–122 (1990).
- Gagosian, R. B., Peltzerand, E. T. & Zafrou, O. C. Atmospheric transport of continentally derived lipids to the tropical North Pacific. *Nature* **291**, 312–315 (1981).
- Gogou, A. I., Aposolaki, M. & Stephanou, E. G. Determination of organic molecular markers in marine aerosols and sediments: one-step flash chromatography compound class fractionation and capillary gas chromatographic analysis. *J. Chromatogr. A* **799**, 215–231 (1998).
- Cavalli, F. *et al.* Advances in identification of organic matter in marine aerosol. *J. Geophys. Res.* **109**, doi:10.1029/2004JD0051377 (2004).
- Gershey, R. M. Characterization of seawater organic matter carried by bubble-generated aerosols. *Limnol. Oceanogr.* **28**, 309–319 (1983).

17. Mochida, M., Kitamori, Y., Kawamura, K., Nojiri, Y. & Suzuki, K. Fatty acids in the marine atmosphere: Factors governing their concentrations and evaluation of organic films on sea salt particles. *J. Geophys. Res.* **107**, doi:10.1029/2001JD001278 (2002).
18. Oppo, C. *et al.* Surfactant component of marine organic matter as agents for biogeochemical fractionation of pollutants transport via marine aerosol. *Mar. Chem.* **63**, 235–253 (1999).
19. Blanchard, D. C. Bubble scavenging and the water to air transfer of organic material in the sea. *Adv. Chem. Ser.* **145**, 360–387 (1976).
20. Geever, M. *et al.* Measurements of primary marine aerosol fluxes at Mace Head, Ireland. In Abstracts of European Aerosol Conference, Madrid, 2003, Vol. 1 *J. Aer. Sci.* S637–S638 (2003).
21. Martensson, E. M., Nilsson, E. D., de Leeuw, G., Cohen, L. H. & Hansson, H. C. Laboratory simulations and parameterization of the primary marine aerosol production. *J. Geophys. Res.* **108**, doi:10.1029/2002JD002263 (2003).
22. Facchini, M. C., Mircea, M., Fuzzi, S. & Charlson, R. J. Cloud albedo enhancement by surface-active organic solutes in growing droplets. *Nature* **401**, 257–259 (1999).
23. Nenes, A. *et al.* Can chemical effects on cloud droplet number rival the first indirect effect? *Geophys. Res. Lett.* **29**, doi: 10.1029/2002GL015295 (2002).
24. Beardall, J. & Raven, J. A. The potential effects of global climate change on microalgal photosynthesis, growth and ecology. *Phycologia* **43**, 26–40 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was partly supported by the European Commission (Projects QUEST and PHOENICS), Irish Research Council for Science, Engineering and Technology, and the Irish Higher Education Authority, Italian Ministry of Environment (Italy–USA Cooperation on Science and Technology of Climate Change). SeaWiFS chlorophyll products were provided by the SeaWiFS project, NASA/Goddard Space Flight Center and ORBIMAGE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C.F. (mc.facchini@isac.cnr.it).

Basal tyrannosauroids from China and evidence for protofeathers in tyrannosauroids

Xing Xu^{1*}, Mark A. Norell², Xuewen Kuang³, Xiaolin Wang¹, Qi Zhao¹ & Chengkai Jia¹

¹Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China

²American Museum of Natural History, New York 10024, USA

³Tianjin Museum of Natural History, Tianjin 300074, China

* Present address: American Museum of Natural History, Central Park West at 79th Street, New York City, New York 10024, USA

Tyrannosauroids are one of the last and the most successful large-bodied predatory dinosaur groups^{1–5}, but their early history remains poorly understood. Here we report a new basal tyrannosauroid from the Early Cretaceous Yixian Formation of western Liaoning, China, which is small and gracile and has relatively long arms with three-fingered hands. The new taxon is the earliest known unquestionable tyrannosauroid found so far^{6–9}. It shows a mosaic of characters, including a derived cranial structure resembling that of derived tyrannosauroids^{1–5} and a primitive postcranial skeleton similar to basal coelurosaurians. One of the specimens also preserves a filamentous integumentary covering similar to that of other coelurosaurian theropods from western Liaoning. This provides the first direct fossil evidence that tyrannosauroids had protofeathers.

Theropoda Marsh, 1881
Coelurosauria *sensu* Gauthier, 1986
Tyrannosauroida Osborn, 1905
Dilong paradoxus gen. et sp. nov.

Etymology. The generic name is derived from Chinese *di*

(emperor) + *long* (dragon). The specific name refers to the surprising characters of this animal.

Holotype. A semi-articulated skeleton including an almost complete skull, IVPP (Institute of Vertebrate Paleontology and Paleoanthropology, Beijing) V14243.

Referred material. IVPP V14242, a nearly complete skull and associated presacral vertebrae; TNP01109 (in collections of Tianjin Museum of Natural History), a partial skull. IVPP V11579 is here referred to this taxon; however, on further analysis it may be determined that it represents a second, closely related species of *Dilong*.

Horizon and locality. Lujiatun, Beipiao, western Liaoning; older than 128 and younger than 139 million years fine sand beds of the lower part of the Yixian Formation¹⁰.

Diagnosis. A small tyrannosauroid distinguishable from other tyrannosauroids by the unique presence of two large pneumatic recesses dorsal to the antorbital fossa on the maxilla, a Y-shaped crest formed by the nasals and lacrimals, an extremely long descending process of the squamosal extending close to the mandibular articulation of the quadrate, a lateral projection of the basisphenoid anterior to the basal tuber, very deep, sub-circular interspinous ligamentous fossae on cervical vertebrae, robust scapula with a wide distal end (distal end twice as wide as the proximal scapular blade) and a hypertrophied coracoid (dorsoventral length about 70% of the scapular length).

Dilong paradoxus is a small tyrannosauroid (see Supplementary Information for ontogenetic assessments and measurements of the specimens). IVPP V14243, the largest of the known specimens, is estimated to be 1.6 m in body length. The small size of *Dilong paradoxus*, in comparison with more derived, large tyrannosauroids, is consistent with the trend of size increase reported for this clade^{7,9,11}. *Dilong paradoxus* shares numerous derived cranial similarities with other tyrannosauroids (Fig. 1a–f)^{1–5,12}. The premaxilla has a deep subnasal body. The anterior ventral edge of the maxilla is convex, and dorsally the maxilla clearly excludes the nasal from participating in the antorbital fossa, a unique feature to tyrannosauroids¹³. In lateral view, a small promaxillary fenestra is visible anteroventral to the maxillary fenestra at the anterior margin of the antorbital fossa. The nasals are fused at an early ontogenetic stage, as in a few theropods and other tyrannosauroids^{1–5,14}, but they are more similar to other tyrannosauroids because of their convex dorsal surface^{1,3}. The pneumatic lacrimal is robust. The descending process of the lacrimal is evidently concave along the orbital edge in lateral view; anteroventrally it floors the antorbital fossa with the maxilla and jugal. The jugal is anteriorly pneumatic and contributes significantly to the border of the antorbital fenestra. Right under the orbit, the ventral margin of the jugal is bent and thickened, forming a well-developed corneal process. The prefrontal is reduced, although it still separates the posterior margin of the lacrimal from the frontal. The supratemporal fossae, separated by a low sagittal crest, extend extensively onto the frontals, occupying large portions of their lateroposterior edges. The lateral nuchal crest is fairly high relative to most other theropods. The quadratojugal is massive and dorsally expanded modestly, suggesting the presence of a weak quadratojugal–squamosal flange dividing the lateral temporal fenestra. The quadrate is posterodorsally orientated and has a pneumatic opening on the posterior surface as in many coelurosaurians¹⁵. The supraoccipital ridge is prominent and the paroccipital process is laterally directed. The basicranium is extensively pneumatized, with a deep basisphenoidal recess that contains large foramina, and is posterovertrally orientated. The mandibular fenestra, if present, must be extremely small. The articular is pneumatic and has a reduced retroarticular process that is orientated posteriorly. A long supradentary is present; however, because of preservation it is not apparent if it was fused to a separate coronoid. As in other tyrannosauroids^{1,3}, the premaxillary teeth are closely packed and more mediolaterally than anteroposteriorly orientated. They are

D-shaped, serrated, and much smaller than the maxillary teeth. The maxillary and dentary teeth (Fig. 1f) are strongly laterally compressed and, as in velociraptorines¹⁶, *Eotyrannus*⁸ and juvenile tyrannosaurids¹³, have posterior serrations that are significantly larger than the anterior ones. Different from most tyrannosaurids, *Dilong paradoxus* has a less robust skull, a much larger external naris, an expanded braincase, a lower sagittal crest, an antero-posteriorly longer basicranium and a smaller mylohyoid foramen, and lacks a surangular foramen (Fig. 1a–e). Some of these features might be size related; however, progressively larger size is a phylogenetic trend in tyrannosaurids^{7,9,11}. It is also different from other tyrannosaurids in several unique features. In lateral view, the maxilla is depressed dorsal to the well-defined dorsal edge of the antorbital fossa. Within this longitudinal depression are two large pneumatic recesses. In dorsal view, a Y-shaped crest is present, which is formed by the nasals and lacrimals. The descending process

of the squamosal is extremely long and ventrally extends approximately to the level of the mandibular articulation of the quadrate. Anterior to the basal tuber, the basisphenoid projects laterally to form a lateral projection, a feature unknown in other tyrannosaurids. These features contribute to the cranial diagnosis for the new taxon.

Postcranially *Dilong paradoxus* has proportionally a longer neck and trunk. The opisthocoelous cervical vertebrae have low neural spines and large pneumatic openings on relatively shallow centra (Fig. 1g, h), the posterior dorsal vertebrae have much larger, relatively long and non-pneumatic amphicoelous centra, transversely narrow neural spines, and strap-like broad transverse processes, and the distal caudal vertebrae lack transverse processes and neural spines. Very deep, subcircular interspinous ligamentous fossae are present on cervical vertebrae. In some theropods such as *Allosaurus*¹⁶, the ligament fossae are well developed but are much shallower and

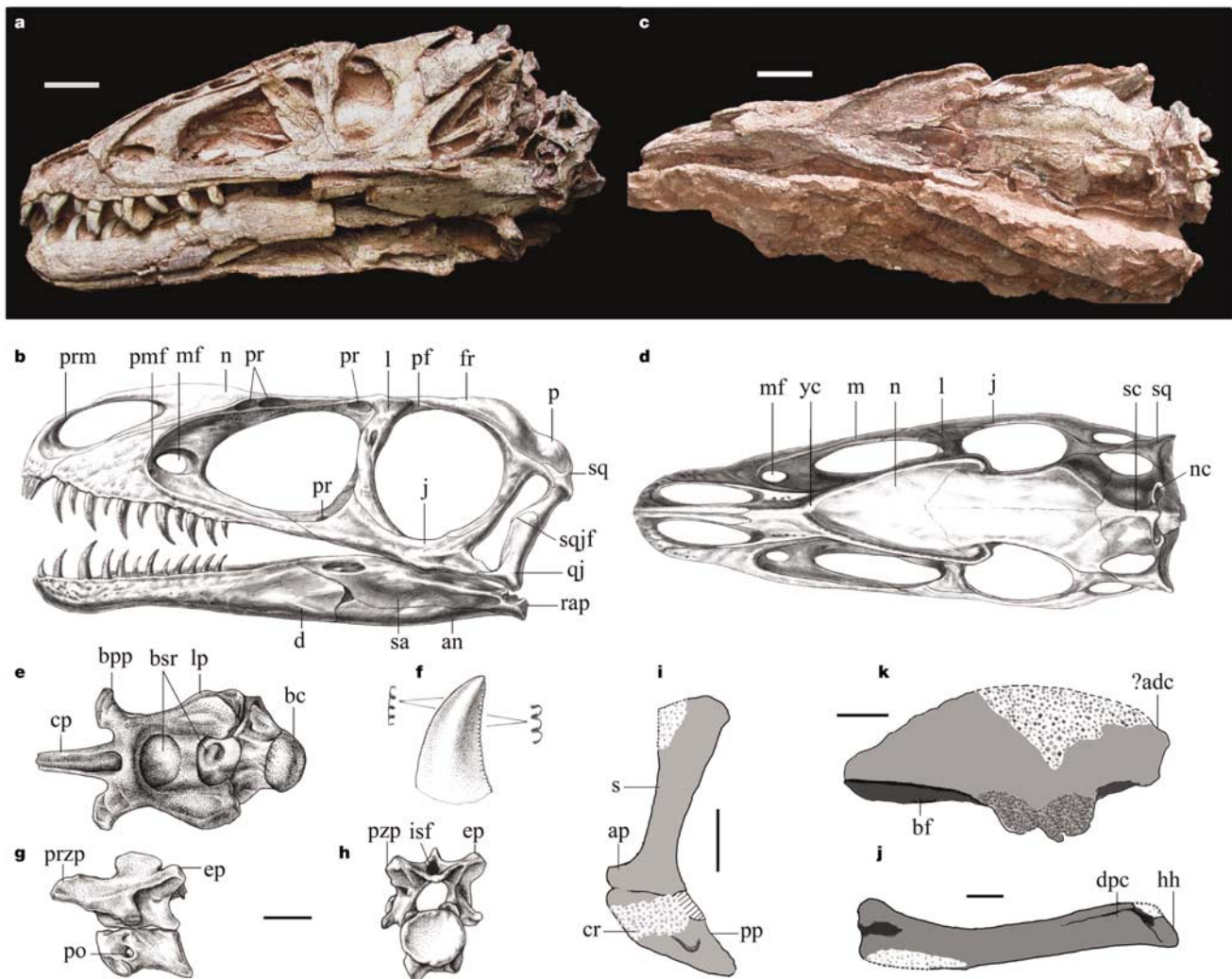


Figure 1 *Dilong paradoxus*. **a, c**, Photographs of the skull of IVPP V14243 in left lateral (**a**) and dorsal (**c**) views. Scale bar, 2 cm. **b, d**, Cranial reconstruction in left lateral (**b**) and dorsal (**d**) views. **e**, Braincase of IVPP V14242 in ventral view. **f**, A dentary tooth of IVPP V14242 in lingual view, with detailed views of the anterior and posterior margins. **g, h**, A middle cervical vertebra of IVPP V14242 in left lateral (**g**) and posterior (**h**) views. Scale bar, 1 cm. **i**, Left scapula and coracoid of IVPP V14243 in lateral view. Scale bar, 2 cm. **j**, Left humerus of IVPP V14243 in anterior view. Scale bar, 1 cm. **k**, Left ilium of IVPP V14243 in medial view. The acetabular region is crushed. Scale bar, 2 cm. Abbreviations: ?adc, ?anterodorsal concavity; an, angular; ap, acromion process; bc, basioccipital

condyle; bf, brevis fossa; bpp, basioptrygoid process; bsr, basisphenoid recess; cp, cultriform process; cr, coracoid; d, dentary; dpc, deltopectoral crest; ep, epiphysis; fr, frontal; hh, humeral head; isf, interspinous ligamentous fossa; j, jugal; l, lacrimal; lp, lateral process; m, maxilla; mf, maxillary fenestra; n, nasal; nc, nuchal crest; p, parietal; pf, prefrontal; pmf, promaxillary fenestra; po, pneumatic opening; pp, posterior process; pr, pneumatic recess; prm, premaxilla; przp, prezygapophysis; pzp, postzygapophysis; qj, quadratojugal; rap, retroarticular process; s, scapula; sa, surangular; sc, sagittal crest; sq, squamosal; sqjf, squamosal–quadratojugal flange; yc, Y-shaped crest.

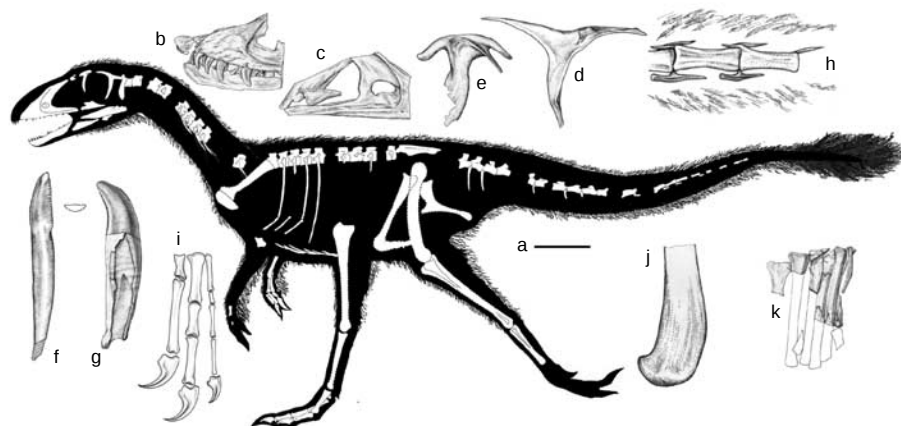


Figure 2 IVPP V11579. **a**, Skeletal reconstruction showing preserved bones. **b**, Left maxilla in lateral view. **c**, Right maxilla in lateral view. **d**, Left postorbital in medial view. **e**, Right squamosal in lateral view. **f**, Premaxillary tooth in lingual view. Note the flat lingual

surface. **g**, Maxillary tooth in labial view. **h**, Distal caudal vertebrae associated with branched integumentary structures. **i**, Reconstructed left manus in dorsal view. **j**, Left ischium in lateral view. **k**, Left metatarsals in anterior view. Scale bar, 10 cm (**a**).

more groove-like. The scapula is robust, with a distal end about twice as wide as the proximal scapular blade. The acromion process of the scapula is dorsally extended significantly. The coracoid is hypertrophied and is about 70% of the scapular length (Fig. 1i). The humerus is more than half of the femoral length and is significantly longer than the scapula, a feature seen in some derived maniraptorans¹⁵. As in other tyrannosauroids⁵, the proximal half of the humerus is narrow, with a proximal end that is only slightly differentiated (Fig. 1j). The distal end of the radius is flattened. The pelvis is less robust than in other tyrannosauroids. The ilium is considerably shorter than the femur and is similar to that of maniraptorans in having a tapering posterior blade and it lacks a broad ventral hook-like projection from the anterior blade (Fig. 1k)^{1,18}. The pubis has an extremely large pubic boot (more than half the pubic length)¹, but with little anterior extension. The hindlimb proportions indicate cursorial capability, with tibia/femur and metatarsus/femur ratios of 1.13 and 0.66, respectively. However, considering the allometric growth in the tyrannosauroid development^{19,20}, it has shorter lower legs than in similar-sized tyrannosauroids. The wing-like lesser trochanter of the femur is significantly lower than the femoral head. As in derived dromaeosaurids¹⁵, the metatarsus is relatively robust. Unlike other tyrannosauroids¹⁹, metatarsal III is proximally unconstricted and weakly distally ginglymoid. Metatarsal V is elongate, though not to the same degree as in dromaeosaurids¹⁵.

IVPP V14242 is smaller than IVPP V14243 and represents an earlier ontogenetic stage than the latter. It differs from IVPP V14243 in having smaller external nares, more posteriorly positioned maxillary fenestrae, a more expanded braincase, a more slender ventral process of the squamosal, a less developed and more posteriorly located lateral process of the basisphenoid and a longer retroarticular process, and in lacking a sagittal crest and a nuchal crest.

A fragmentary specimen (IVPP V11579), which was collected from approximately 125-million-year-old grey shale of the Yixian Formation¹⁰ of Zhangjiagou locality, Beipiao, western Liaoning, can be referred to *Dilong* on the basis of the identical morphologies of most overlapping elements, including squamosals, dentaries, splenials, teeth, cervical vertebrae, dorsal vertebrae, caudal vertebrae, and metatarsals (Fig. 2). IVPP V11579 provides additional significant information unavailable in the known specimens of *Dilong paradoxus*. The preserved manual elements show a robust manual digit II where metacarpal II and associated phalanges are much more robust than metacarpal I and corresponding phalanges. This is a derived feature also seen in other tyrannosauroids⁵. Manual digit

III, however, is not reduced in phalangeal number or digit length, though extremely slender, representing a precursor to the reduced manual digit III in more advanced tyrannosauroids¹. The relatively long arms with three-fingered hands in Liaoning tyrannosauroids conform to predictions of current phylogenetic hypotheses about the morphology of basal tyrannosauroids^{7,21}. In contrast with the derived tyrannosauroids¹ but similar to many basal coelurosaurians²¹, the distal end of the ischium is slightly expanded and curved anteriorly.

On the tail, traces of filamentous integumentary structures are preserved at an angle of about 30–45° to the caudal vertebral series (Fig. 3). A small patch of filamentous integumentary structures is also preserved close to the posterior left mandible. Those attached to the distal caudal vertebrae are more than 20 mm in length. They are branched but the pattern is difficult to discern. They seem to be composed of a series of filaments joined at their bases along a central filament as in *Sinornithosaurus*²². Many of these structures seem to be distally branched, as can be observed in some cases in *Sinornithosaurus* (IVPP V12811).

The fossil record of the tyrannosauroids is mainly restricted to the Late Cretaceous of North America and central and eastern Asia^{1–5}. Several Early Cretaceous and Late Jurassic taxa have been reported, but they are either represented by limited material or are questionably referable^{6–9}. Liaoning tyrannosauroids represent the most basal tyrannosauroid species (see Supplementary Information for a numerical cladistic analysis) represented by multiple, nearly complete specimens found so far and provide new insights into the early evolution of tyrannosauroid dinosaurs.

Comparatively, Liaoning tyrannosauroids are similar to other juvenile tyrannosauroids; for example, they both have long and low proportions of snout and mandible, teeth with differently sized anterior and posterior serrations, low cervical neural spines, relatively low cervical and dorsal centra, and dorsal centra much larger than cervical ones^{13,23}. Some of these features have also been reported in other basal tyrannosauroids⁸. These data indicate that peramorphosis might be important in tyrannosauroid evolution, pending an analysis on a more complete data set.

Liaoning tyrannosauroids show more of the derived modifications of tyrannosauroids in the cranium than postcranium. This might also be so in *Eotyrannus*⁸. For example, Liaoning tyrannosauroids have 80% cranial diagnostic features relative to 25% postcranial synapomorphies of tyrannosauroids (see Supplementary Information)^{1–4}. This indicates that the cranial modification occurred earlier than the postcranial modification in tyrannosauroid evolution.

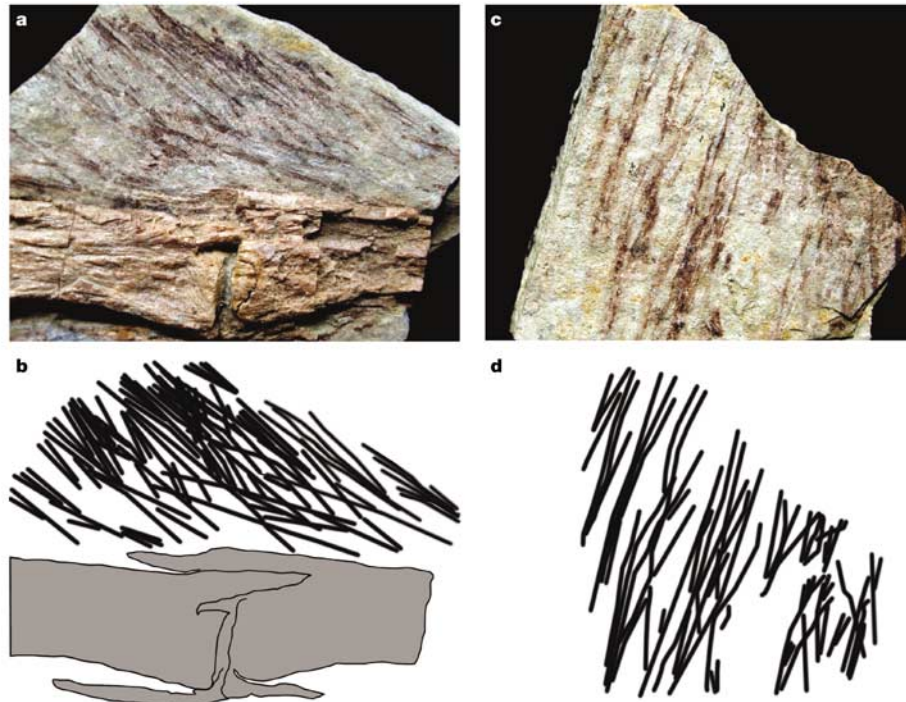


Figure 3 Integumentary structures of IVPP V11579. **a, b**, Filamentous integumentary structures along the dorsal edge of the distal caudal vertebrae, photograph (**a**) and

linedrawing (**b**). **c, d**, Close-up of the integumentary structures showing the simple branching pattern, photograph (**c**) and linedrawing (**d**). Not to scale.

Several features are significant for understanding coelurosaurian evolution. In comparison with the gigantic, derived Late Cretaceous tyrannosauroids, Liaoning tyrannosauroids have a less pneumatic skeleton. Postcranial pneumatization is, as in tyrannosauroids, also less developed in small, basal members of many non-avian coelurosaurian groups^{15,24} and more developed in large, derived members of these groups. The distribution of postcranial pneumatization is thus very complex among coelurosaurians^{15,24}, rather than displaying a continuous evolutionary trend along the line to birds. Liaoning tyrannosauroids lack an arctometatarsalian pes, a feature often suggested to be associated with a strong cursorial capability¹⁹. The arctometatarsalian pes has been proposed to diagnose a more inclusive group, Arctometatarsalia²⁵. However, recent discoveries show that this interesting pedal feature, as in tyrannosauroids, is absent from small, basal members of most 'arctometatarsalian' subgroups^{15,26}, and is developed in the derived, larger members of the respective groups. In other taxa such as alvarezsaurids and troodontids, the arctometatarsalian pes is present even in small-bodied forms. This strongly supports the hypothesis that the arctometatarsalian pes evolved independently within different coelurosaurian groups^{21,27}.

The filamentous integumentary structures in Jehol theropods have been interpreted as protofeathers²². The presence of the similar structures in IVPP V11579 provides the first direct evidence showing that tyrannosauroids possessed protofeathers. Furthermore, the filamentous protofeathers are branched as in other coelurosaurians²². This is a distinctive morphological feature of modern feathers, suggesting that this important modification occurred early in coelurosaurian evolution. Large, derived tyrannosauroids were reported to have scaled skin²⁸, but the presence of two kinds of body covering is not unexpected. However, current understanding of the integumentary morphology in non-avian theropods is hindered by poor information on distribution. Given the diverse morphologies of integumentary structures in living birds, it is possible that non-avian theropods had different integumentary morphologies on different regions of the body, and derived, large

tyrannosauroids might bear both scale-like and filamentous integumentary appendages. Alternatively, the lack of filamentous integumentary structures in derived tyrannosauroids is correlated with the large size, a physiological strategy also adopted by some mammals such as elephants, which lose most of their body hairs as they mature²⁹. This therefore supports the hypothesis that the original function of protofeathers is correlated with thermoregulation³⁰. □

Received 18 May; accepted 13 July 2004; doi:10.1038/nature02855.

- Holtz, T. R. Jr in *Mesozoic Vertebrate Life* (eds Carpenter, K. & Tanke, D.) 64–83 (Indiana University Press, Bloomington, 2001).
- Currie, P. J., Hurum, J. H. & Sabath, K. Skull structure and evolution in tyrannosaurid dinosaurs. *Acta Palaeontol. Pol.* **48**, 227–234 (2003).
- Currie, P. J. Cranial anatomy of tyrannosaurid dinosaurs from the Late Cretaceous of Alberta, Canada. *Acta Palaeontol. Pol.* **48**, 191–226 (2003).
- Hurum, J. H. & Sabath, K. Giant theropod dinosaurs from Asia and North America: skulls of *Tarbosaurus bataar* and *Tyrannosaurus rex* compared. *Acta Palaeontol. Pol.* **48**, 161–190 (2003).
- Brochu, C. A. Osteology of *Tyrannosaurus rex*: insights from a nearly complete skeleton and high-resolution computed tomographic analysis of the skull. *J. Vertebr. Paleontol. Mem.* **7**, 1–138 (2003).
- Buffetaut, E., Suteethorn, V. & Tong, H. The earliest known tyrannosaur from the Lower Cretaceous of Thailand. *Nature* **381**, 689–691 (1996).
- Manabe, M. The early evolution of Tyrannosauridae in Asia. *J. Paleontol.* **73**, 1176–1178 (1999).
- Hutt, S., Naish, D., Martill, D. M., Barker, M. J. & Newbery, P. A preliminary account of a new tyrannosaurid theropod from the Wessex Formation (Early Cretaceous) of southern England. *Cretaceous Res.* **22**, 227–242 (2001).
- Rauhut, O. W. M. A tyrannosauroid dinosaur from the Upper Jurassic of Portugal. *Palaeontology* **46**, 903–910 (2003).
- Swisher, C. et al. Further support for a Cretaceous age for the feathered-dinosaur beds of Liaoning, China: new ⁴⁰Ar/³⁹Ar dating of the Yixian and Tuchengzi formations. *Chin. Sci. Bull.* **46**, 2009–2013 (2001).
- Erickson, G. M. et al. Gigantism and comparative life history of *Tyrannosaurus rex*. *Nature* **430**, 772–775 (2004).
- Molnar, R. E. The cranial morphology of *Tyrannosaurus rex*. *Palaeontographica A* **217**, 137–176 (1991).
- Currie, P. J. & Dong, Z. New information on *Shanshanosaurus huoyanshanensis*, a juvenile tyrannosaurid (Theropoda, Dinosauria) from the Late Cretaceous of China. *Can. J. Earth Sci.* **38**, 1729–1737 (2001).
- Sampson, S. D. et al. Predatory dinosaur remains from Madagascar: implications for the Cretaceous biogeography of Gondwana. *Science* **280**, 1048–1051 (1998).
- Xu, X. *Deinonychosaurian Fossils from the Jehol Group of Western Liaoning and the Coelurosaurian Evolution* Thesis, Chinese Academy of Sciences, Beijing (2002).
- Currie, P. J. New information on the anatomy and relationships of *Dromaeosaurus albertensis* (Dinosauria: Theropoda). *J. Vertebr. Paleontol.* **15**, 576–591 (1995).
- Madsen, J. H. *Allosaurus fragilis*: a revised osteology. *Utah Geol. Min. Surv. Bull.* **109**, 1–163 (1976).

18. Gauthier, J. in *The Origin of Birds and the Evolution of Flight* (ed. Padian, K.) 1–55 (California Academy of Sciences, San Francisco, 1986).
19. Holtz, T. R. The arctometatarsalian pes, an unusual structure of the metatarsus of Cretaceous Theropoda (Dinosauria: Saurischia). *J. Vertebr. Paleontol.* **14**, 480–519 (1995).
20. Currie, P. J. Allometric growth in tyrannosaurids (Dinosauria: Theropoda) from the Upper Cretaceous of North America and Asia. *Can. J. Earth Sci.* **40**, 651–665 (2003).
21. Holtz, T. R. Jr A new phylogeny of the carnivorous dinosaurs. *Gaia* **15**, 5–61 (2000).
22. Xu, X., Zhou, H. H. & Prum, R. O. Branched integumental structures in *Sinornithosaurus* and the origin of feathers. *Nature* **410**, 200–204 (2001).
23. Carr, T. D. Craniofacial ontogeny in Tyrannosauridae (Dinosauria, Theropoda). *J. Vertebr. Paleontol.* **19**, 497–520 (1999).
24. Hwang, S. H., Norell, M. A., Ji, Q. & Gao, K. New specimens of *Microraptor zhaoianus* (Theropoda: Dromaeosauridae) from northeastern China. *Am. Mus. Novit.* **3381**, 1–44 (2002).
25. Holtz, T. R. The phylogenetic position of the Tyrannosauridae: implications for the theropod systematics. *J. Paleontol.* **68**, 1100–1117 (1994).
26. Xu, X., Norell, M. A., Wang, X. L., Makovicky, P. J. & Wu, X. C. A basal troodontid from the Early Cretaceous of China. *Nature* **415**, 780–784 (2002).
27. Sereno, P. C. The evolution of dinosaurs. *Science* **284**, 2137–2147 (1999).
28. Martin, L. D. & Czerkas, S. A. The fossil record of feather evolution in the Mesozoic. *Am. Zool.* **40**, 687–694 (2000).
29. Spinge, C. *Elephants* (T. & A.D. Poyser, London, 1994).
30. Chen, P. J., Dong, Z. M. & Zhen, S. N. An exceptionally well-preserved theropod dinosaur from the Yixian Formation of China. *Nature* **391**, 147–152 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Holtz, O. Rauhut and X.-C. Wu for critical comments on the manuscript; Z. H. Zhou, Z. L. Tang, Y. Q. Wang, Y. Li, H. J. Wang, Y. L. Huo, H. Q. Shou, X.Z. Liu, Q. Cao, W. Chen, J. C. Lu and C. Li for their contribution in the field; H. J. Wang, J. M. Yang, G. H. Cui and X. Q. Ding for preparing the specimens; and R. S. Li, J. L. Huang and M. W. Yang for illustrations. The study was supported by the Special Funds for Major State Basic Research Projects of China, the National Natural Science Foundation of China, the National Geographic Society, the Chinese Academy of Sciences, the National Science Foundation of the USA and the American Museum of Natural History.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.X. (xu@amnh.org or xingxu@vip.sina.com) or M.A.N. (norell@amnh.org).

Pleistocene to Holocene extinction dynamics in giant deer and woolly mammoth

A. J. Stuart¹, P. A. Kosintsev², T. F. G. Higham³ & A. M. Lister¹

¹Department of Biology, University College London, London WC1E 6BT, UK

²Institute of Plant and Animal Ecology, 620144 Ekaterinburg, Russia

³Oxford Radiocarbon Accelerator Unit, University of Oxford, Oxford OX1 3JQ, UK

The extinction of the many well-known large mammals (mega-fauna) of the Late Pleistocene epoch has usually been attributed to 'overkill' by human hunters, climatic/vegetational changes or to a combination of both^{1,2}. An accurate knowledge of the geography and chronology of these extinctions is crucial for testing these hypotheses. Previous assumptions that the mega-fauna of northern Eurasia had disappeared by the Pleistocene/Holocene transition² were first challenged a decade ago by the discovery that the latest woolly mammoths on Wrangel Island, northeastern Siberia, were contemporaneous with ancient Egyptian civilization^{3,4}. Here we show that another spectacular mega-faunal species, the giant deer or 'Irish elk', survived to around 6,900 radiocarbon yr BP (about 7,700 yr ago) in western Siberia—more than three millennia later than its previously accepted terminal date^{2,5}—and therefore, that the reasons for its ultimate demise are to be sought in Holocene not Pleistocene events. Before their extinction, both giant deer and woolly mammoth

underwent dramatic shifts in distribution, driven largely by climatic/vegetational changes. Their differing responses reflect major differences in ecology.

The giant deer, *Megaloceros giganteus* Blumenbach, with a maximum shoulder height of 2.1 m and an antler span of up to 3.6 m, is one of the most striking and evocative extinct animals of the Palaeolithic. First appearing about 0.4 Myr ago⁶, the overall distribution of *M. giganteus* during the Last Cold Stage extended across the middle latitudes of Eurasia from Ireland to east of Lake Baikal⁷. Its anatomy and distribution suggest it was a mixed feeder, requiring both to browse and to graze in a productive environment—especially necessary to sustain the annual antler growth in males^{6,8}. On the other hand, it is likely that the huge antlers would have excluded males from even moderately dense woodland, at least for part of the year.

The giant deer was a key element in the relatively restricted set of Late Pleistocene extinctions in northern Eurasia, but unlike the mammoth it was not a species of the treeless 'steppe-tundra'. Theories of its extinction formerly invoked the 'maladaptation'⁹, or more recently the seasonal nutrient requirements⁸, of the huge antlers, and have focused on the well-studied Irish population of the Allerød phase around 12–10.6 kyr (12,000 to 10,600 uncalibrated radiocarbon yr BP; all dates hereafter are radiocarbon yr BP, unless stated otherwise. See Supplementary Information for calibrated calendar equivalents). Ireland has yielded most of the best-preserved specimens, including near-complete skeletons, from calcareous lake sediments ideal for the preservation of bone^{10,11}. The absence of specimens from deposits of the succeeding Younger Dryas phase (about 10.6–10 kyr) has led to an extrapolated assumption of global extinction at the onset of this severe cold episode^{5,9,10}.

Fortunately, most Late Quaternary extinctions occurred well within the range of radiocarbon (¹⁴C) dating, and we use direct dating of megafaunal remains, thereby minimizing any uncertainties of stratigraphical context. So far we have obtained 43 new radiocarbon accelerator mass spectrometry (AMS) dates from the Oxford Radiocarbon Accelerator Unit (ORAU) directly from *M. giganteus* material from western Europe, the Urals and Siberia, to which have been added 49 previously published direct dates (ORAU and other laboratories) (see Supplementary Information). This direct-dating approach has allowed us to track the fate of the species through climatic episodes leading up to extinction across its geographic range and in relation to changing palaeoenvironments.

During the Last Cold Stage, until around 20 kyr, the giant deer was widespread across western Europe (Fig. 1a), although the records are sporadic, and to the north of the Mediterranean region its occurrence may have been restricted to warmer (interstadial) phases with suitable vegetational growth. Indirectly dated records suggest that the species was probably also present in southeastern Europe and central Asia¹² during this time, and there are two dates around 39 kyr and 41 kyr (close to or beyond the reliable limit of ¹⁴C dating) from Kamenka Buryatia, Transbaikalia (Supplementary Information). However, there are no known occurrences before 20 kyr from the Urals and the adjacent part of western Siberia.

Our data so far indicate a striking absence of dates for giant deer within the long interval around 20–12.5 kyr, implying that it had withdrawn entirely from western and central Europe. This period corresponds broadly to the Last Glacial Maximum (LGM), when the Scandinavian and Alpine ice sheets expanded, and elsewhere open treeless steppe-tundra vegetation predominated. Indirectly dated records¹² suggest that refugia for *Megaloceros* at this time (Fig. 1b) included parts of southeastern Europe and south central Asia, probably in areas where tree and shrub vegetation persisted. So far there are no records for giant deer, ¹⁴C-dated or otherwise, from Mediterranean Europe for the LGM period, although in view of the known persistence of woodland¹³ (Supplementary Information) this region presumably would have provided suitable refugia for this animal.

Many dates and abundant secure stratigraphic evidence^{5,10,14–16} show that after the LGM, *Megaloceros* re-occupied part of north-western Europe. The earliest good evidence that we have for this re-occupation (discounting an old date with a large error from Britain) is a date of $12,455 \pm 65$ (Oxford laboratory number OxA-11687) from the Isle of Man, as well as a series of dates clearly indicating extensive recolonization of the Isle of Man and Britain by around 12.2 kyr, and of Ireland, southern Scandinavia and northern Germany by about 12 kyr (Figs 1c and 2). Whether or not these apparent differences in timing are real, needs to be tested by further dating. This Late Glacial recolonization by giant deer and its persistence for some one and a half millennia (Figs 1d and 2) can be correlated with the Late Glacial Interstadial (LGI), a warm phase that began around 13–12.6 kyr. The rather rapid warming at the onset of the LGI resulted in the replacement of steppe–tundra with a more productive grass and sedge vegetation with shrubs and some birch trees^{17,18} (Supplementary Information).

However, there is no evidence that giant deer returned to central or southern parts of western Europe. There are very few published accounts of *Megaloceros* from putatively Late Glacial contexts in these areas. Most of the specimens that we were able to trace proved to have been misidentified, and the only two that are definitely giant deer did not give a ¹⁴C date.

The apparent restriction of *Megaloceros* within Europe to part of the northwest in the LGI (Fig. 1c, d) is intriguing. In the generally cooler, later part of the LGI (Allerød), birch and pine woodland colonized much of Europe, whereas more open herbaceous veg-

etation with birch trees was present in the north-west; therefore, the distribution of giant deer at this time plausibly correlates to a marked north–south vegetational gradient (Supplementary Information). But this cannot explain the same distribution pattern in the early LGI when similar open vegetation occurred across both northwestern and central Europe. A recent analysis of the ¹⁴C-dated archaeological record¹⁹ indicates that after a major contraction in range during the LGM, humans progressively repopulated central and northwestern Europe from the south, and had reached northern Germany and Britain—albeit in low population densities—by around 14 kyr, well before the reappearance of giant deer in this region. Therefore there is no indication of human interference with giant deer recolonization of northwestern Europe in the Late Glacial. However, it is possible that denser human populations further south, shown by significantly higher numbers of occupation sites (C. Gamble and W. Davies personal communication), may have inhibited recolonization of central and southern Europe by *Megaloceros* after the LGM.

The subsequent dramatic collapse of the northwestern European population of *Megaloceros*, which occurred around 10.7 kyr, inferred both from the pattern of ¹⁴C dates and abundant stratigraphic evidence of its absence from Younger Dryas sediments^{5,10,11,14,15}, can be correlated with the marked deterioration in vegetation in this cold phase, when open steppe–tundra returned¹⁸ (Supplementary Information). However, dates of $10,585 \pm 65$ (OxA-11498) and $10,257 \pm 75$ (Arizona laboratory number, AA-51350) (Fig. 2) on an antler from the River Cree, southwestern

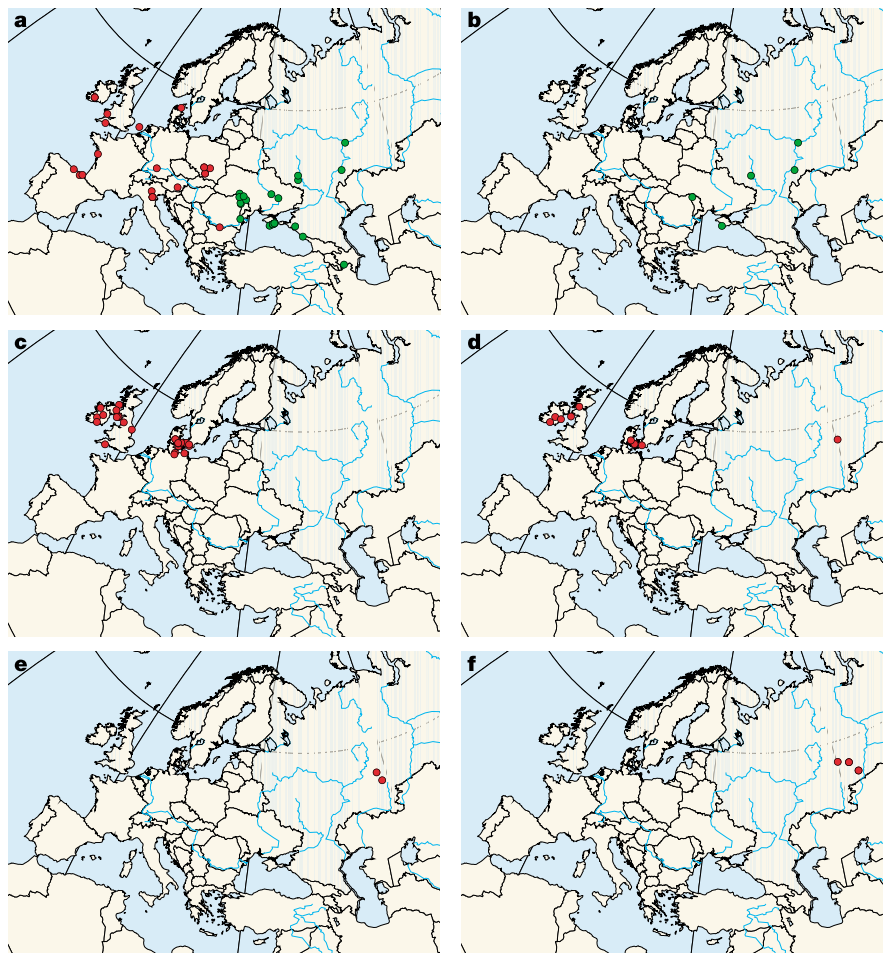


Figure 1 Radiocarbon dated (red dots) *M. giganteus* finds, showing chronological changes in distribution. Green dots indicate indirectly dated records for eastern Europe¹². Dates are uncalibrated. **a**, 20 kyr and older. **b**, Around 20–13 kyr, indicating possible

refugia during the LGM. **c**, 12.9–11 kyr. **d**, 10.9–10.5 kyr. **e**, 10.4–9.9 kyr. **f**, 8.0–6.9 kyr Shigir (Urals), Kamyshev, Redut (western Siberia).

Scotland, may indicate persistence into the early part of the Younger Dryas in northwestern Britain. In both Ireland and the Isle of Man (probably connected to mainland Britain) extirpation was almost certainly due solely to climatic/vegetational change because there is no evidence of humans from these areas until the early Holocene epoch, around 9 kyr (ref. 20).

Claims of Holocene survival in northwestern Europe have not been substantiated. Two recent ^{14}C dates obtained directly on antlers seemed to demonstrate that giant deer had persisted through the Younger Dryas and into the earliest Holocene in the Isle of Man and southwestern Scotland²¹, but further investigation indicates that these specimens actually date from the Late Glacial (see Methods and Supplementary Information). We have been unable to locate *Megaloceros* remains found by workmen at Theresienhof (northern Germany) and said to have come from a Younger Dryas/Preboreal horizon²². Suggestions of Holocene survival in Ireland²³ are based on specimens of very dubious provenance¹⁰.

However, a series of new AMS dates presented here indicates that giant deer did persist long into the Holocene in a very different area, on the boundary between Europe and Asia in the Urals/western Siberia region. The earliest record that we have from this region is about 10.8 kyr, during the late Allerød or early Younger Dryas (Figs 1d and 2; Table 1). The appearance of giant deer here probably represents colonization from southeastern Europe or west central Asia, rather than from northwestern Europe, which is considerably further away. A date from Kulmetovsk cave (Table 1; Figs 1e and 2), on the eastern side of the southern central Urals, falls within the Younger Dryas; another, from Grotto Bobylek, is close to the transition from the Younger Dryas to the Holocene. This is followed by a date of around 8 kyr from the Shigir Mesolithic site (Fig. 1f). The two youngest records of *Megaloceros* (around 6.9 and 7.0 kyr) known so far from anywhere are, respectively, on a largely complete

antler-bearing male skeleton from Kamyshlov mire (Supplementary Fig. 1) and a male skull with associated cervical vertebrae from the Redut river, in the adjacent part of western Siberia. These were each independently AMS dated by the Kiel (KIA) and Oxford laboratories, and the dates are statistically indistinguishable (Table 1 and Fig. 2). The validity of Holocene survival is supported not only by records from two separate sites and replicate results from two laboratories, but also by the fact that they are linked to their Late Pleistocene predecessors by a series of intermediate dates.

There are two crucial questions. First, why did *Megaloceros* survive through the Younger Dryas in the Urals and western parts of Siberia, but not in western Europe? Palaeobotanical data indicate that in contrast to the harsh conditions of the Younger Dryas in Europe, on the eastern slopes of the Urals this period saw the persistence of grass-shrub vegetation and open woodland with larch, spruce, pine and birch trees²⁴—apparently good habitat for *Megaloceros*—whereas less suitable, dry steppe-tundra with dwarf birch and dwarf willow occupied the adjacent West Siberian Plain. Subsequently, in the Early Holocene, the spread of closed forest in the mountains probably excluded giant deer, but it survived in the forest-steppe of western Siberia²⁴. In view of the limited available data for large areas of its known distribution, it is possible that further work will reveal other areas in eastern Europe and Asia (for example, southern Siberia) where *Megaloceros* survived into the Holocene. A single new *Megaloceros* date of $10,055 \pm 45$ (OxA-13026) from the Chernigovo open cast coal mine, Kuznetsk basin in southwestern Siberia²⁵, suggests that the giant deer persisted through the Younger Dryas in this region, raising the possibility that it may also have survived into the Holocene.

The second important question is what caused its extinction in the Holocene? Given the comparative rarity of *Megaloceros* material, the Redut and Kamyshlov individuals may not have been among the

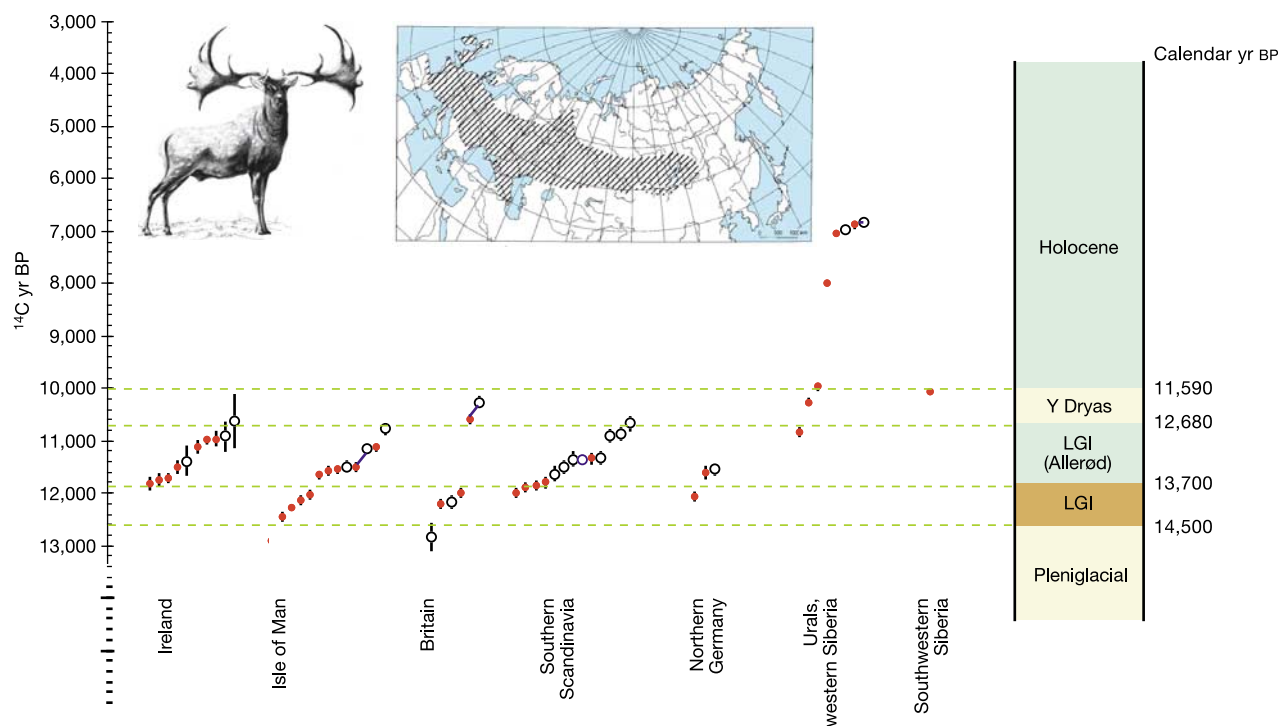


Figure 2 Chart of Late Glacial and Holocene ^{14}C dates for *M. giganteus*. Inset: overall Last Cold Stage distribution⁷. Error bars are ± 1 s.d. Connected points are dates made on the same material. Oxford AMS dates are shown in red (Supplementary Information), other laboratories are open circles^{5,14–16,21}. Calendar (varve) dates are indicated for major

boundaries¹⁸. Major climatic/vegetational phases are shown schematically^{17,18} (Supplementary Information). LGI, Late Glacial Interstadial; Y Dryas, Younger Dryas. Predominant vegetation: steppe-tundra, yellow; shrub vegetation, brown; trees and shrubs, green.

Table 1 ¹⁴C-dated *Megaloceros* finds from the Urals (UR) and western Siberia (WS)

Locality	Co-ordinates	Provenance	Element	Laboratory number	¹⁴ C date	δ ¹³ C†
Kamyshlov mire WS	62.73° E 56.92° N	Peat bog	Rib (associated skeleton)	KIA-5669	6816 ± 35	−21.46
Kamyshlov mire WS	62.73° E 56.92° N	Peat bog	Skull (associated skeleton)	OxA-13015	6881 ± 38	−19.5
Redut*, Miass river WS	64.23° E 55.48° N	Peat bog	Cervical vertebra (skull and vertebrae)	KIA-5668	6968 ± 33	−20.16
Redut, Miass river WS	64.23° E 55.48° N	Peat bog	Cervical vertebra (skull and vertebrae)	OxA-13014	7034 ± 34	−19.0
Shigir WS	60.17° E 57.37° N	Mesolithic site	Antler	OxA-11064	7990 ± 45	−19.7
Grotto Bobylek UR	57.65° E 56.32° N	Cave	Tooth root	OxA-11063	9960 ± 55	−19.6
Kulmetovsk UR	58.47° E 55.15° N	Cave	Maxilla	OxA-10676	10260 ± 55	−19.5
Neviansk WS	60.20° E 57.50° N	Alluvium	Skull	OxA-11065	10825 ± 65	−19.7

* The duplicate dates from Kamyshlov and Redut are in each case from the same individual.

† δ¹³C = [(¹³C/¹²C)_{sample} / (¹³C/¹²C)_{standard}] − 1 × 1,000

last of their line, even in the Urals/western Siberia region; however, the spread of dry steppe over western Siberia around 7 kyr while closed forest persisted in the mountains²⁴ may have resulted in or contributed to the extirpation of this population.

A full evaluation of the possible role of humans in the processes of range reduction and ultimate extinction will require further data from across the Siberian part of the range of *Megaloceros*, both in terms of radiocarbon chronology and the archaeological record. In the Urals/western Siberia region, known Mesolithic and Neolithic settlements are largely located on the banks of rivers and lakes²⁶ (Supplementary Information). It is possible that while *Megaloceros* survived in the Ural foothills it was relatively safe from human predation, but when forced onto the plain by vegetational changes, it became critically vulnerable to increased hunting pressure.

Previous studies have suggested that the nutrient requirements of antler growth proved critical at times of environmental stress, and that the latest (Allerød) populations from the westernmost margin of the species' range may have suffered reduction in body and antler size^{8,11,21}. However, the Kamyshlov skeleton—a large male with an antler span of 2.56 m (ref. 27; Supplementary Fig. 1)—shows no evidence of reduction in antler or body size. Although it is only a single specimen, its antler, skull and limb bone dimensions fall in the middle of the range of variation of the Irish Late Glacial sample, itself composed largely of males^{6,9,11,27}. The Redut specimen, a skull with neck vertebrae, was unfortunately not available for measurement.

The striking contrast in the histories of giant deer and woolly mammoth, *Mammuthus primigenius* Blumenbach, can be plausibly linked to the differing ecologies of the two species. *Megaloceros* shows a much larger apparent LGM gap in Europe, around 20–12.5 kyr compared with 18–16 kyr for mammoth (Fig. 3). Giant deer did not re-expand until the onset of the LGI when increased warmth probably resulted in increased plant productivity¹⁷, and it apparently became particularly abundant when open habitats were invaded by trees and shrubs in the Allerød. In contrast, only about 0.5 kyr after the reappearance of *Megaloceros* in northwestern Europe, mammoth disappeared from most of northern Eurasia, and available data indicate that after 12 kyr—when forests spread at the onset of the Allerød interstadial—it became restricted to the far north of Siberia where open steppe–tundra environments persisted³ (Supplementary Information). The latest available mainland records (Taymyr Peninsula) are around 9.6 kyr (ref. 3; Supplementary Information). A series of ¹⁴C dates, from more than one laboratory, demonstrates that mammoth survived to around 3.7 kyr (about 4,000 yr ago) on Wrangel Island, northeastern Siberia, where steppe–tundra vegetation occurs at the present day⁴ (Supplementary Information), and recently it has been shown that it survived to around 8 kyr on St Paul Island (Pribilofs)

in the Bering Sea²⁸. The contrasting patterns of giant deer compared with woolly mammoth (Fig. 4) highlight the profound impact of climate and vegetation on the range fluctuations of these two species before their extinction. Their complementary chronological distribution also strongly suggests that the pattern of presence and absence in each is not due to taphonomic or other extraneous effects.

The complex history of the last approximately 30 kyr of woolly mammoth and giant deer (Irish elk), emphasizes the 'ragged' nature of the so-called 'end Pleistocene' megafaunal extinctions in northern Eurasia, in which different species went extinct at different times, and had their 'last stands' in different regions². Our perspective on the extinction of both species has shifted radically with the realization that we have to consider Holocene as well as Late Pleistocene events in seeking the cause or causes of their demise, with the important corollary that there were no major climatic/vegetational changes in the Holocene compared with those of the Last Cold Stage. Widely distributed before about 20 kyr, both giant deer and woolly mammoth underwent successive contractions in

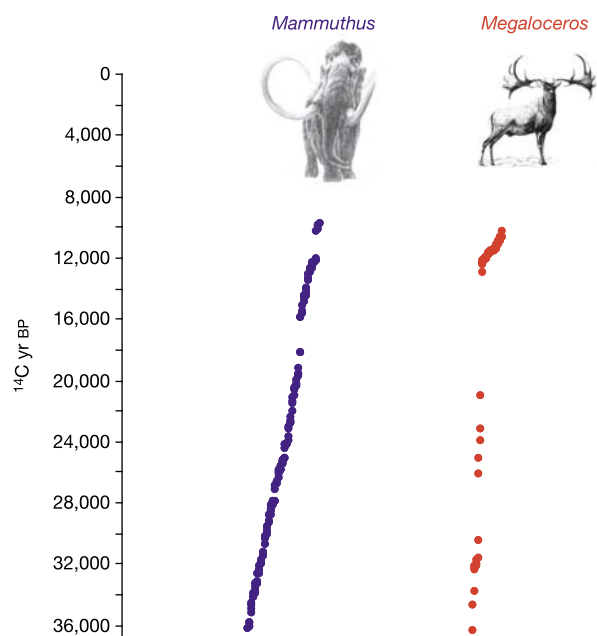


Figure 3 Chart of ¹⁴C dates <36 kyr for *Mammuthus primigenius* and *Megaloceros giganteus* in western Europe. Note much larger LGM gap for giant deer than for mammoth (refs as Fig. 2, and refs 3, 29; Supplementary Information).

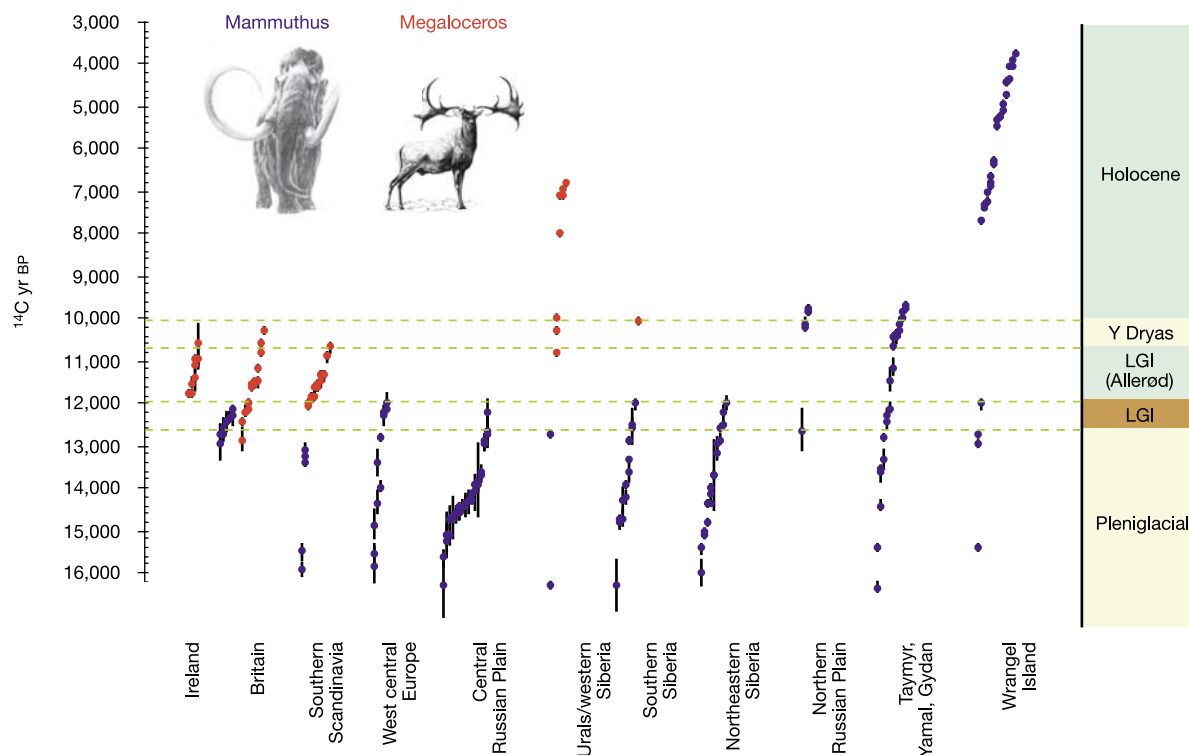


Figure 4 Chart of ^{14}C dates <16 kyr for *Megaloceros giganteus* (red) and *Mammuthus primigenius* (blue) in Europe and selected areas of Siberia^{3,4,9–13,29} (and refs as Fig. 2). Error bars are ± 1 s.d. Climatic/vegetational phases as in Fig. 2. For clarity, Isle of Man dates are grouped with Britain, and northern Germany with southern Scandinavia. Note the increased records for giant deer in Europe around 12–11 kyr, after the disappearance

of mammoth. Note also the termination of most mammoth dates around 12 kyr (Allerød), except for parts of Siberia, and outlying mammoth dates around 10 kyr in northeastern Europe, probably representing brief repopulation from northern Siberia during the Younger Dryas^{3,29}.

range (although at different times), from each of which they made only partial recoveries. Finally, in the Holocene, both were confined to one or more relatively small areas before becoming extinct. Comparison of our data set with proxy records indicates that the marked shifts in the distributions of both species were driven by climate acting through vegetational changes. However, giant deer had survived previous Pleistocene cold episodes, and unlike the mammoth—which presumably survived previous interglacials in Siberian refugia—had expanded its European range and thrived during previous interglacials. Why, then, didn't giant deer re-expand its range in the Holocene as in previous interglacials with putatively similar climate and vegetation¹³? The possible role of modern humans (not present in previous interglacials) needs further investigation. Future work will need to target other possible Holocene refugia, and further analyse the spatio-temporal pattern of extinction in comparison with vegetational change and human activity. □

Methods

Sample selection

Careful attention was given to specimen identification in view of the frequent past misidentification of red deer, elk/moose or large bovids such as *Megaloceros*. The direct-dating approach is crucial in avoiding stratigraphical uncertainty, as is replicate dating of key samples.

Radiocarbon dating

Bone samples were prepared for AMS ^{14}C dating at the ORAU using routine collagen extraction procedures (Supplementary Information). An additional ultrafiltration pre-treatment step was used to further purify the bone gelatin and retain only the >30-kD molecular weight fraction for ^{14}C assay. Lyophilized gelatin samples were combusted using a Roboprep CHN sample converter unit and mass spectrometrically analysed using a Europa Scientific 20-20 mass spectrometer, operating in continuous flow mode. Graphite was prepared using routine methods (Supplementary Information). Bone collagen

preservation was evaluated using the carbon to nitrogen atomic ratio (C:N) and the per cent weight collagen was extracted from the bone. The addition of exogenous carbon atoms increases the C:N ratio and, depending upon the age and size of the contaminant, may result in errors in the AMS determinations. All of the *Megaloceros* bones dated at ORAU were within the acceptable C:N range of 2.9–3.5. All dated samples exceeded 1% wt collagen, which is the minimum threshold for acceptance.

When dating duplicate samples of bone, ORAU AMS determinations often yielded slightly older results compared with other laboratories (Fig. 2 and Table 1). We attribute this to the removal of low molecular weight (<30 kD) components, including potential contaminants, by ultrafiltration. These particles include degraded and broken up collagen, salts, sediment particulates and sometimes contaminants of different, usually more modern, ^{14}C age. Low-collagen-yield bones (pre-treatment yield < approximately 20 mg) must be treated with care during ultrafiltration to ensure that humectants (glycerin) designed to keep the ultrafilters moist are removed completely before use. Failure to do this has resulted in small amounts (approximately 20–30 μg) of glycerin becoming incorporated into gelatin, affecting accuracy in dating low-collagen-yield bones of Holocene age, because the glycerin is 30 kyr in age. ORAU therefore implements a rigorous cleaning protocol of all ultrafilters before use, to ensure removal to background levels (Supplementary Information). None of the dated samples was of low enough pre-treatment yield to be affected by this contaminant.

The young Redut and Kamyshlov specimens were dated both at the Leibniz Labour AMS laboratory (Kiel) and in Oxford (Table 1). The Kiel bone samples were prepared to filtered collagen using a similar method to that used at ORAU, but without ultrafiltration. As a test of the presence of potential contaminants, both filtered collagen and insoluble residue fractions from both bones were dated. The age difference between the KIA-5668 residue and collagen suggests that the bone was not significantly contaminated with exogenous carbon. Similarly, the results are identical to the date of OxA-13014, obtained from the same bone. The Kamyshlov sample (KIA-5669) yielded a much younger age for the removed insoluble residue, suggesting the bone has been contaminated, but that this has been effectively removed by the pre-treatment chemistry applied to the bone. The ORAU ultrafiltered AMS result was again statistically indistinguishable from the Kiel filtered collagen date. In a previous study²¹, remains of *M. giganteus* from Ballaugh, Isle of Man and the River Cree, southwestern Scotland, were dated to the early Holocene (around 9.4–9.2 kyr). Re-dating of duplicate samples at ORAU by the above methods (Supplementary Information), and separately by the University of Arizona NSF Radiocarbon Laboratory, have produced ages within the Late Glacial, and these ages are accepted here (Fig. 2 and Supplementary Information).

Received 26 August 2003; accepted 27 July 2004; doi:10.1038/nature02890.

- Martin, P. S. in *Quaternary Extinctions: A Prehistoric Revolution* (eds Martin, P. S. & Klein, R. G.) 364–403 (Univ. Arizona Press, Tucson, 1984).
- Stuart, A. J. Mammalian extinctions in the late Pleistocene of northern Eurasia and North America. *Biol. Rev.* **66**, 453–562 (1991).
- Stuart, A. J., Sulerzhitsky, L. D., Orlova, L. A., Kuzmin, Y. V. & Lister, A. M. The latest woolly mammoths (*Mammuthus primigenius*) in Europe and Asia: a review of the current evidence. *Quat. Sci. Rev.* **21**, 1559–1569 (2002).
- Vartanyan, S. L., Garrut, V. E. & Sher, A. V. Holocene dwarf mammoths from Wrangel Island in the Siberian Arctic. *Nature* **382**, 337–340 (1993).
- Barnosky, A. D. Big game extinction caused by late Pleistocene climatic change: Irish elk (*Megaloceros giganteus*) in Ireland. *Quat. Res.* **25**, 128–135 (1986).
- Lister, A. M. The evolution of the giant deer, *Megaloceros giganteus* (Blumenbach). *Zool. J. Linn. Soc.* **112**, 65–100 (1994).
- Kahlke, R. D. Die Entstehungs-, Entwicklungs- und Verbreitungsgeschichte des Oberpleistozänen *Mammuthus-Coelodonta*-Faunenkomplexes in Eurasien (Großsäuger). *Abh. Senckenb. Natf. Ges.* **546**, 1–64 (1994).
- Moen, R. A., Pastor, J. & Cohen, Y. Antler growth and extinction of Irish Elk. *Evol. Ecol. Res.* **1**, 235–249 (1999).
- Gould, S. J. The origin and function of 'bizarre' structures: antler size and skull size in the 'Irish Elk', *Megaloceros giganteus*. *Evolution* **28**, 191–220 (1974).
- Mitchell, G. F. & Parkes, H. M. The giant deer in Ireland. *Proc. R. Ir. Acad. B* **52**, 291–314 (1949).
- Barnosky, A. D. Taphonomy and herd structure of the extinct Irish Elk *Megaloceros giganteus*. *Science* **228**, 340–344 (1985).
- Markova, A. K. et al. Late Pleistocene distribution and diversity of mammals in northern Eurasia. *Paleont. Evol.* **28–29**, 5–143 (1995).
- Tzedakis, P. C. & Bennett, K. D. Inter-glacial vegetation succession: a view from southern Europe. *Quat. Sci. Rev.* **14**, 967–982 (1996).
- Kaiser, K., De Klerk, P. & Terberger, T. Die "Riesenhirsch Fundstelle" von Endingen: geowissenschaftliche und archäologische Untersuchungen an einem spätglazialen Fundplatz in Vorpommern. *Eiszeit. Ggw.* **49**, 102–122 (1999).
- Aaris-Sørensen, K. & Liljegren, R. Late Pleistocene remains of giant deer (*Megaloceros giganteus* Blumenbach) in Scandinavia – chronology and environment. *Boreas* **33**, 61–73 (2004).
- Woodman, P., McCarthy, M. & Monaghan, N. The Irish Quaternary fauna project. *Quat. Sci. Rev.* **16**, 129–159 (1997).
- Hoek, W. Z. Vegetation response to the 14.7 and 11.5 ka cal. BP climate transitions: is vegetation lagging climate? *Global Planet. Change* **30**, 103–115 (2001).
- Litt, T., Schmincke, H.-U. & Kromer, B. Environmental response to climatic and volcanic events in central Europe during the Weichselian Lateglacial. *Quat. Sci. Rev.* **22**, 7–32 (2003).
- Gamble, C., Davies, W., Paul Pettitt, P. & Richards, M. Climate change and evolving human diversity in Europe during the last glacial. *Phil. Trans. R. Soc. Lond. B* **359**, 243–254 (2004).
- Woodman, P. C. in *The Quaternary History of Ireland* (eds Edwards, K. & Warren, W.) 251–278 (Academic, Orlando and London, 1985).
- Gonzalez, S., Kitchener, A. C. & Lister, A. M. Survival of the Irish Elk into the Holocene. *Nature* **405**, 753–754 (2000).
- Guenther, E. W. Funde von Ren und Riesenhirsch in nacheiszeitlichen Schichten aus der Nähe von Theresenhof bei Plön. *Schrift. Naturwiss. Ver. Schleswig-Holstein* **30**, 79–81 (1960).
- Vereshchagin, N. K. & Baryshnikov, G. F. in *Quaternary Extinctions: A Prehistoric Revolution* (eds Martin, P. S. & Klein, R. G.) 483–516 (Univ. Arizona Press, Tucson, 1984).
- Panova, N. K. *The Forest Development Process in Urals and Trans-Urals* [in Russian] 26–49 (Uralskoye Otdelenie Rossiiskoi Akademii Nauk, Ekaterinburg, 1996).
- Foronova, I. V. *Quaternary Mammals of the South-East of Western Siberia (Kuznetsk Basin): Phylogeny, Biostratigraphy, and Palaeoecology* [In Russian] 243 (United Institute of Geology Transactions 848, Publishing house of SB RAS Branch "GEO", Russian Academy of Sciences, Siberian Branch, Novosibirsk, 2001).
- Serikov, Y. B. *Paleolithic and Mesolithic of the Middle Trans-Urals* [in Russian] 351 (Izdatelstvo Nizhne-Tagil'skogo Pedagogicheskogo Instituta, Nizhny Tagil, 2000).
- Pavlova, M. V. Etudes sur l'Histoire paléontologique des Ongulés, IX Sélénodontes post tertiaires de la Russie. *Mem. Acad. Sci. St Petersburg* **20**, 1–95 (1906).
- Guthrie, R. D. Radiocarbon dating evidence of Holocene mammoths on an Alaskan Bering Sea island. *Nature* **429**, 746–748 (2004).
- Stuart, A. J. The extinction of woolly mammoth (*Mammuthus primigenius*) and straight-tusked elephant (*Palafoxodon antiquus*) in Europe. *Quat. Int.* (in the press).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The research was funded by the Natural Environment Research Council. We thank the following for their help in providing specimens and/or data: K. Aaris-Sørensen, J. Altuna, N. Benecke, P. Boscato, M. Breda, C. Bronk Ramsey, V. Cabrera, A. Currant, W. Davies, P. Doughty, I. V. Foronova, C. Gamble, M. Germonpré, S. Gonzalez, K. Hawkins, R. Hedges, B. Huntley, R. Jacobi, K. James, L. Kaagan, R. Kahlke, A. Kitchener, Y. V. Kuzmin, J. van der Made, A. K. Markova, D. Mol, D. Nagel, L. Niven, M. Palombo, M. Patou, P. Pettitt, A. Pinto, B. Sala, M. Street, A. Tagliacozzo, P. Tomlinson, E. Turner, A. A. Vorobiev and P. Wojtal. Special thanks to I. V. Foronova for agreeing to publication of the Chernigovo date in advance of further collaborative work on Siberian *Megaloceros*, and to P. Grootes for AMS radiocarbon dating of the Kamyshlov and Redut specimens. We are grateful to A. Mangione, B. Sala and Superintendenza per i Beni Archeologici dell'Umbria for permission to use the *Megaloceros* illustration in Figs 2, 3 and 4.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.J.S. (Tony@megafauna.freemove.co.uk).

Photosynthetic architecture differs in coastal and oceanic diatoms

Robert F. Strzepek^{1*} & Paul J. Harrison^{2*}

¹Department of Botany, University of British Columbia, Vancouver BC Canada V6T 1Z4

²Department of Botany and Department of Earth and Ocean Sciences, University of British Columbia, Vancouver BC Canada V6T 1Z4

* Present addresses: NIWA Centre for Chemical and Physical Oceanography, Department of Chemistry, University of Otago, PO Box 56, Dunedin, New Zealand (R.F.S.); Atmospheric, Marine & Coastal Environment (AMCE) Program, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong (P.J.H.)

Diatoms are a key taxon of eukaryotic phytoplankton and a major contributor to global carbon fixation¹. They are ubiquitous in the marine ecosystem despite marked gradients in environmental properties, such as dissolved iron concentrations, between coastal and oceanic waters. Previous studies have shown that offshore species of diatoms and other eukaryotic algae have evolved lower iron requirements to subsist in iron-poor oceanic waters, but the biochemical mechanisms responsible for their decreased iron demand are unknown^{2,3}. Here we show, using laboratory-cultured model species, a fundamental difference between a coastal and an oceanic diatom in their photosynthetic architecture. Specifically, the oceanic diatom had up to fivefold lower photosystem I and up to sevenfold lower cytochrome *b₆f* complex concentrations than a coastal diatom. These changes to the photosynthetic apparatus markedly decrease the cellular iron requirements of the oceanic diatom but not its photosynthetic rates. However, oceanic diatoms might have also sacrificed their ability to acclimate to rapid fluctuations in light intensity—a characteristic of dynamic and turbid coastal waters. We suggest that diatoms, and probably other eukaryotic algal taxa, exploited this difference in the underwater light climate between oceanic and coastal waters, enabling them to decrease their iron requirements without compromising photosynthetic capacity. This adaptation probably facilitated the colonization of the open ocean by diatoms, and contributes to their persistence in this iron-impooverished environment.

It has been well established that diatoms and other eukaryotic algae from oceanic waters are less susceptible to iron limitation than coastal species^{2–6}. This trend was supported by growth rate data from two closely related model diatom species cultured over an ecologically relevant range of iron concentrations and light intensities (Fig. 1). In iron-replete medium, the two diatoms grew at comparable rates (Fig. 1a). However, the oceanic centric diatom, *Thalassiosira oceanica*, maintained high growth rates (about 80% of those in iron-replete conditions) in low-iron media that restricted the growth of the coastal diatom, *Thalassiosira weissflogii*, to about 20% of iron-replete growth rates (Fig. 1c). Previous studies have shown that oceanic species of diatoms and other eukaryotic algae are unusual in that they maintain these high growth rates despite having very low cellular iron concentrations^{2,3,6}. We therefore conducted the first investigations at the biochemical level to determine how an oceanic diatom has decreased its dependence on iron.

Theoretical calculations predict that most of the iron required by phytoplankton is used for photosynthetic electron transport. Iron is an essential component of the cytochrome and iron-sulphur protein cofactors of the major photosynthetic complexes: photosystem II (PSII), the cytochrome (Cyt) *b₆f* complex and photosystem I (PSI)^{7,8}. Although the iron content of the complexes is evolutionarily conserved, there is some flexibility in iron requirements because the cellular abundance of the complexes themselves

differs between algal taxa^{7–9} and change in response to growth irradiance and iron availability^{7,8,10}. The relative abundance of the complexes is also a key determinant of phytoplankton iron requirements because the Cyt *b₆f* complex (six atoms of iron per complex) and PSI (12 atoms of iron per complex) are disproportionately iron-rich compared with PSII (two atoms of iron per complex). Our measurements of these complexes revealed that the biochemical composition of the oceanic isolate is strikingly different from any other photosynthetic organism yet examined. The oceanic diatom contained concentrations of PSII comparable to those in the coastal diatom (Fig. 2a,b), but its concentrations of Cyt *b₆f* complex (Fig. 2c,d) and PSI (Fig. 2e,f) were unusually low, even in iron-replete cells. Consequently, the PSII:PSI ratio of *T. oceanica* was consistently about 10:1 (Fig. 2a,e). In contrast, the PSII:PSI ratio of coastal diatoms (and other Chl *c*-containing algae) is generally about 2:1 (Fig. 2b,f); in land plants the ratio is about 1:1 (ref. 8). Cytochrome *f* concentrations (the proxy for the Cyt *b₆f* complex) were also substantially lower in *T. oceanica* (3.7–9.0 $\mu\text{mol l}^{-1}$) than in *T. weissflogii* (7.5–31 $\mu\text{mol l}^{-1}$) and other coastal diatoms^{11,12}.

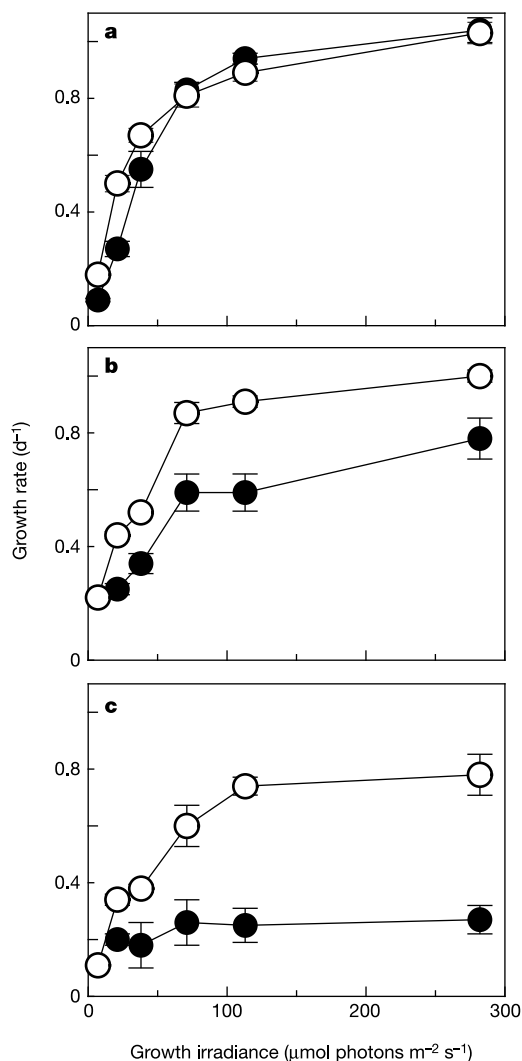


Figure 1 Steady-state growth rates of the oceanic diatom *T. oceanica* and the coastal diatom *T. weissflogii* acclimated to a range of irradiances and iron concentrations. **a**, Iron-replete medium (total iron (Fe_T) = 59 nM); **b**, **c**, low-iron media (Fe_T = 7.8 nM (**b**) and 4.7 nM (**c**)). Open symbols, *T. oceanica*; filled symbols, *T. weissflogii*. Error bars indicate s.e.m. ($n \geq 6$).

Because PSI and the Cyt *b₆f* complex contain considerably more iron than PSII^{7,8}, these biochemical alterations are indicative of an adaptive strategy employed by oceanic diatoms to minimize their iron requirements.

Using the biochemical data (Fig. 2), we estimated the proportion of cellular iron associated with photosynthetic electron transport (see Methods). Our calculations indicate that in iron-limited diatoms, most cellular iron (50–100%) is contained within the photosynthetic electron carriers (Fig. 3c,d), in excellent agreement with theoretical predictions⁷. Consequently, changes in the quantity of iron contained within the photosynthetic apparatus due to photoacclimation were reflected in the bulk cellular iron content of the diatoms ($R^2 = 0.81$ for *T. weissflogii*; $R^2 = 0.93$ for *T. oceanica*; $P < 0.05$), providing empirical support for previous inferences that low-light-acclimated diatoms have higher cellular iron concentrations because they contain more photosynthetic electron carriers¹⁰. Most significantly, our results provide a novel explanation for the different iron requirements of oceanic and coastal diatoms. When iron-replete, the oceanic isolate had twofold to threefold less iron within the photosynthetic apparatus than the coastal isolate (Fig. 3a,b), chiefly owing to its much lower concentration of Cyt *b₆f* complex and the absence of Cyt *c₆* (Fig. 2c,d). That this interspecific difference in photosynthetic iron requirements exists in iron-replete diatoms (but cannot be inferred directly from cellular iron concentrations) is particularly arresting, because it implies that *T. oceanica* might not acclimate to low iron availability as such, but rather has evolved an intrinsically lower iron requirement.

The remarkable feature of *T. oceanica* is not so much its lower

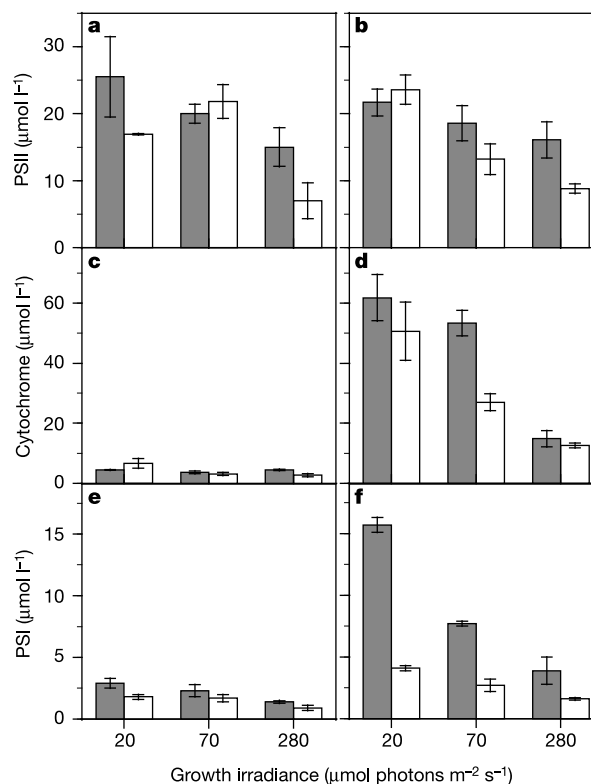


Figure 2 Photosystem and electron carrier concentrations of *T. oceanica* and *T. weissflogii*. **a**, **b**, PSII; **c**, **d**, Cyt *f* and Cyt *c₆* (cytochrome); **e**, **f**, the reaction centre of PSI, P700. Filled bars, iron-replete (Fe_T = 59 nM) cultures; open bars, low-iron (Fe_T = 7.8 nM) cultures. **a**, **c**, **e**, *T. oceanica*; **b**, **d**, **f**, *T. weissflogii*. The Cyt *f*: Cyt *c₆* ratio in *T. weissflogii* was about 1:1; Cyt *c₆* was not detected in *T. oceanica*. Error bars indicate s.e.m. ($n \geq 3$).

cellular iron concentrations² but that it can maintain growth (Fig. 1) and photosynthetic rates (Table 1) comparable to those of the coastal diatom despite its high PSII:PSI ratio and low concentration of Cyt *b*₆*f*. In contrast, the loss of PSI in iron-limited cells of *T. weissflogii* (Fig. 2f), which also resulted in lower cellular iron concentrations (Fig. 3d), was concurrent with lower growth rates (Fig. 1). This result implies that the effective absorption cross-section and turnover rate of PSI must be considerably greater than for PSII in the oceanic diatom. For example, if PSII absorbed more excitation energy than PSI, the intersystem electron carriers (that is, the plastoquinone pool) would become too reduced, thereby impairing photosynthetic efficiency. Our results demonstrate that the oceanic diatom can effectively balance excitation energy and electron flow when grown under low light intensity, as the high photochemical quenching observed ($q_p = 0.76$; Table 1) is indicative of limited plastoquinone reduction, and the maximum quantum yield of photosynthesis (as inferred from the initial slope of the photosynthesis–irradiance curve (α^B)) was substantially greater in the oceanic diatom than in the coastal diatom (Table 1). More striking differences in the photosynthetic physiology of the diatom isolates were evident in cells acclimated to high light intensity.

At high irradiances, iron-replete cultures of *T. oceanica* exhibited a substantial decrease in F_v/F_m (about 45%; Table 1). Kinetic experiments indicated that this decrease was due to both the ‘static’ quenching and photoinhibition of PSII reaction centres (data not shown). The irradiance-dependent decline in F_v/F_m occurred at much lower irradiances ($40 \mu\text{mol photons m}^{-2} \text{s}^{-1}$; about 3% of the ocean surface irradiance on a clear day) than observed in coastal diatom isolates (more than $1,000 \mu\text{mol photons m}^{-2} \text{s}^{-1}$) (ref. 13). However, the photosynthetic capacity (P_{max}) of *T. oceanica* was not

affected by PSII photoinhibition within our experimental irradiance range: P_{max} was apparently maintained by means of a compensatory twofold increase in steady-state electron transfer ($1/\tau$) through the remaining functional PSII reaction centres (Table 1). Oceanic phytoplankton assemblages often display concurrent midday maxima in photosynthetic rates and minima in F_v/F_m , indicating that this compensatory mechanism might also operate at higher irradiances encountered *in situ*¹⁴.

Photoinhibition of PSII (but not photosynthesis) seemed to result from the decreased capacity of *T. oceanica* to downregulate light harvesting effectively at even relatively low irradiances (about $40 \mu\text{mol photons m}^{-2} \text{s}^{-1}$). Chlorophyll quenching data indicated that *T. oceanica* was particularly prone to PSII photoinhibition because of its limited ability to dissipate excess excitation energy through non-photochemical quenching (NPQ = 0.06; Table 1). Kinetic analyses revealed that *T. oceanica* was specifically deficient in short-term, ‘energy-dependent’ fluorescence quenching (data not shown) that works by switching PSII antennae into a state of thermal dissipation instead of efficient light harvesting. This photoprotective mechanism—which is regulated by the energization (the ΔpH) of the thylakoid membrane produced chiefly through proton translocation by the Cyt *b*₆*f* complex^{15–17}—seems to be particularly effective at attenuating the damaging effects of large, rapid (time-scale of minutes) fluctuations in irradiance^{18,19}. Such a light regime is characteristic of coastal waters, which are both turbid and subject to rapid tidal stirring. In contrast, similar variations in light intensity occur over hours to days in the more physically stable and transparent open ocean²⁰. We propose that this fundamental difference between marine habitats in light climate decreased the selective pressure on ancient diatoms, which palaeoceanographic data suggest to have originated in coastal habitats²¹, to retain iron-costly short-term photoprotective mechanisms. They could therefore decrease the amount of iron needed within photosynthetic electron carriers without associated decreases in photosynthetic competence. We suggest that this adaptation facilitated the colonization and enduring inhabitation of the open ocean by diatoms.

Our experiments show that the iron requirements of diatoms, and probably those of other eukaryotic algae, are determined largely by their underlying photosynthetic architecture. Accordingly, we predict that many oceanic eukaryotic algae with low requirements for iron also share a novel photosynthetic physiology comparable to that of our model oceanic diatom: specifically, a diminished capacity for the rapid regulation of light harvesting. Although this attribute might limit their success in dynamic coastal light regimes, it seems to have been a successful compromise that contributed to the progression of eukaryotic phytoplankton from iron-rich coastal environments to the iron-poor open ocean, where the foremost

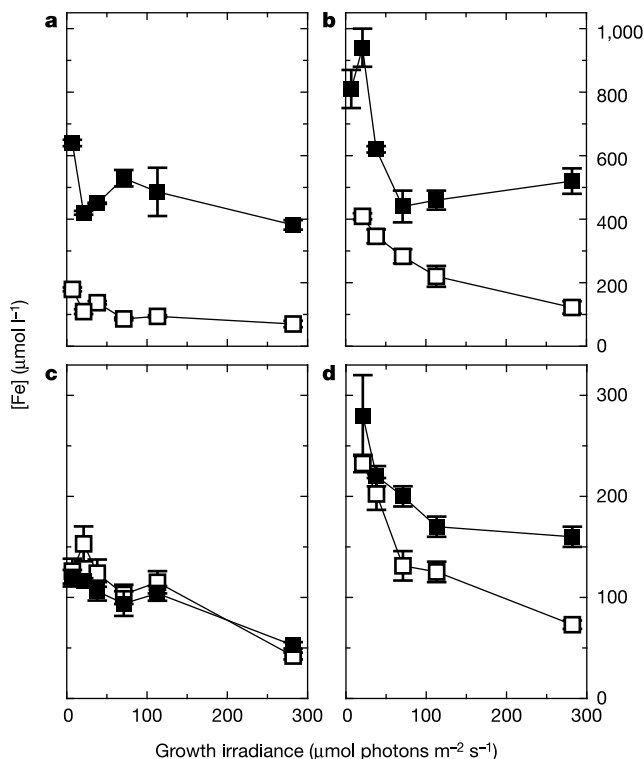


Figure 3 Intracellular and photosynthetic iron concentrations of *T. oceanica* and *T. weissflogii*. **a, c**, *T. oceanica*; **b, d**, *T. weissflogii*. **a, b**, iron-replete ($\text{Fe}_T = 59 \text{ nM}$) medium; **c, d**, low-iron ($\text{Fe}_T = 7.8 \text{ nM}$) medium. Filled symbols, intracellular iron concentration; open symbols, photosynthetic iron concentration. Photosynthetic iron concentrations were calculated as described in Methods. Error bars indicate s.e.m. ($n \geq 3$).

Table 1 Photosynthetic characteristics of iron-replete diatom isolates

Measure	<i>T. oceanica</i>		<i>T. weissflogii</i>	
	Low light	High light	Low light	High light
PSII:Cyt <i>b</i> ₆ <i>f</i> :Cyt <i>c</i> ₆ :PSI	8.7:1.1:0:1	10.5:3.2:0:1	2.2:2.0:1.5:1	4.4:1.9:1.5:1
P_{max}	1.6 ± 0.1	1.8 ± 0.2	1.4 ± 0.1	1.5 ± 0.2
α^B	5.6 ± 0.4	3.8 ± 0.4	1.6 ± 0.2	0.48 ± 0.02
$1/\tau$	0.05 ± 0.01	0.12 ± 0.01	0.08 ± 0.01	0.10 ± 0.02
F_v/F_m	0.69 ± 0.01	0.42 ± 0.02	0.72 ± 0.01	0.64 ± 0.01
q_p	0.76 ± 0.03	0.54 ± 0.01	0.59 ± 0.03	0.82 ± 0.02
NPQ	0.06 ± 0.01	0.06 ± 0.03	0.12 ± 0.01	1.12 ± 0.06

Photosystem and electron carrier molar ratios are reported relative to PSI. Low-light measures were determined for cultures acclimated to a growth irradiance of $20 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ and, for photochemical (q_p) and non-photochemical (NPQ) quenching, the irradiance at which the measurements were performed; high light was $280 \mu\text{mol photons m}^{-2} \text{s}^{-1}$. F_v/F_m is the maximum photochemical efficiency of PSII. Units for maximum photosynthetic rate (P_{max}) are mol O_2 per litre of cell volume per hour. Units for the initial slope of the photosynthesis–irradiance curve normalized to Chl *a* (α^B) are $(\text{mol O}_2 \text{ mol}^{-1} \text{ Chl a h}^{-1})/(\mu\text{mol photons m}^{-2} \text{s}^{-1})$. Units for the steady-state electron transport rate from water to carbon dioxide ($1/\tau$) are ms^{-1} . Results are means $\pm$ s.e.m. ($n \geq 3$).

selective pressure on phytoplankton is to minimize their iron requirements. Our results thus imply that iron availability, through its effect on photoacclimatory strategy, has shaped the evolution of marine phytoplankton and currently influences their geographical distribution. □

Methods

Culturing

The centric diatoms *Thalassiosira oceanica* (isolated from the Sargasso Sea, clone 1003) and *T. weissflogii* (isolated from an estuary on Long Island, New York, USA, clone ACTIN) were grown in semi-continuous batch cultures in trace-metal-buffered (10 µM EDTA) synthetic seawater medium (300 µM nitrate) using trace metal-clean techniques⁶. The basal medium contained 2.1 nM iron contamination, which was determined electrochemically. Iron-replete conditions were obtained by adding 56.5 nM premixed, filter-sterilized FeEDTA (1:1), which, because of the photolability of FeEDTA complexes, corresponds to a range of measured inorganic iron concentrations ($[Fe'] = 385\text{--}714\text{ pM}$ at light intensities of $7\text{--}280\text{ }\mu\text{mol photons m}^{-2}\text{ s}^{-1}$) as determined with the sulphoxine method²². Media containing low concentrations of iron were prepared by adding either 5.7 or 2.6 nM iron ($[Fe'] = 87\text{--}183\text{ pM}$ and $53\text{--}111\text{ pM}$, respectively). Calculated free concentrations of trace metal ions were as follows: Cu, $10^{-13.79}\text{ M}$; Mn, $10^{-8.27}\text{ M}$; Zn, $10^{-10.88}\text{ M}$; Co, $10^{-10.88}\text{ M}$. At each iron concentration, cultures were grown at $18 \pm 1^\circ\text{C}$ in continuous light at six intensities (7, 20, 40, 70, 110, and $280\text{ }\mu\text{mol photons m}^{-2}\text{ s}^{-1}$) obtained with fluorescent bulbs (40 W Vita-Lite; Duro-Test Corp., Philadelphia, Pennsylvania) attenuated with neutral-density screening, and measured with a 4π quantum meter (model QSL-100; Biospherical Instruments, San Diego, California). Specific growth rates (d^{-1}) were calculated from least-squares regressions of natural logarithm of *in vivo* fluorescence against time during the exponential growth phase of acclimated cultures.

Biochemical composition

Replicate ($n \geq 3$) 1.5-litre cultures were harvested during mid-exponential phase. Culture density (cells ml^{-1}) and cell volumes (fl per cell; $1\text{ fl} = 10^{-15}\text{ l}$) were measured with a calibrated Coulter Counter (Model Z2). Biochemical and photosynthetic data were normalized to cell volume because the diatom isolates differed in size and because *T. weissflogii* shrank to about 50% of its maximum size at the lowest combination of iron concentration and growth irradiance examined (see Supplementary Table 1). Oxygen flash yields, cellular concentrations of Cyt *f* and P700 were used as proxies for PSII, the Cyt *b₆f* complex and PSI, respectively. All assays measured functional components. For these calculations, we assumed that Cyt *f* and P700 were present within functional complexes only and in a fixed stoichiometry to the other subunits of their respective complexes. The reaction centre of PSI, P700, was measured by spectroscopy as described previously²³. Reduced versus oxidized spectra were measured in the soluble and membrane fractions for the determination of Cyt *c₆* and Cyt *f*, respectively^{24,25}. The cellular concentrations of functional photosynthetic complexes were calculated from the Chl *a*:complex ratio and cellular Chl *a* concentrations (see Supplementary Table 1), which were measured fluorimetrically. By multiplying the cellular concentration of each complex by its known iron content (2 Fe per PSII, 6 Fe per Cyt *b₆f*, 1 Fe per Cyt *c₆* and 12 Fe per PSI)^{7,8,26}, the iron necessary for linear photosynthetic electron transport was calculated. This calculation does not include ferredoxin (Fd) that, together with flavodoxin (Flv), was measured semiquantitatively by immunoassay. Both diatom species expressed only Fd when grown in iron-replete medium. Ferredoxin was partly replaced by Flv in low-iron ($Fe_T = 7.8\text{ nM}$) cultures of both *T. oceanica* (Fd index = $Fd/(Fd + Flv) = 0.37 \pm 0.16$; mean $\pm$ s.e.m. of all growth irradiance treatments) and *T. weissflogii* (Fd index = 0.44 ± 0.02). Intracellular iron concentrations were measured in parallel cultures grown in 28-mL polycarbonate tubes using the radiotracer $^{55}\text{FeCl}_3$ ($920\text{--}2,015\text{ MBq l}^{-1}$; Amersham) and a Ti(III) EDTA-citrate reducing solution to remove extracellular iron.

Photosynthetic measurements

Photosynthetic measurements (oxygen flash yields (*n*), photosynthesis versus irradiance (*P*–*E*) curves, and pulse-amplitude-modulated (PAM) fluorimetry) were performed on dark-acclimated samples (20 min). Cell concentrates were prepared for *P*–*E* and *n* measurements by gently filtering cells on $1.0\text{-}\mu\text{m}$ (*T. oceanica*) or $3.0\text{-}\mu\text{m}$ (*T. weissflogii*) polycarbonate membrane filters (Poretics, Livermore, California) and resuspending the cells in growth medium sterilized with a $0.2\text{-}\mu\text{m}$ filter. *P*–*E* curves were determined at 18°C with a calibrated Hansatech oxygen electrode system (Model DW1), a quartz–halogen light source and neutral-density filters. Changes in dissolved oxygen concentration were measured in 13 steps from 0 to $2,660\text{ }\mu\text{mol photons m}^{-2}\text{ s}^{-1}$, and a hyperbolic tangent function²⁷ was fitted to the *P*–*E* data. Oxygen flash yields were performed to determine cellular PSII concentrations as described previously²³, except that a single xenon flash lamp (Oriol Corporation, Stratford, Connecticut) was used to provide saturating flashes. The maximum rate at which electrons can be transferred from water to carbon dioxide in the steady state ($1/\tau$; ms^{-1}) was calculated from $1/\tau = P_{\text{max}}^B/n$, where P_{max}^B is the maximum photosynthetic rate normalized to Chl *a* (ref. 28). Fluorescence induction measurements were performed on diatom suspensions by using a PAM 101/103 fluorimeter (Heinz Walz, Effeltrich, Germany) and an emitter–detector–cuvette assembly (ED101). The maximum quantum efficiency of PSII (F_v/F_m) was determined by inducing a transient closure of PSII reaction centres with a light pulse (700 ms , $3,000\text{ }\mu\text{mol photons m}^{-2}\text{ s}^{-1}$) generated with a filtered (Schott KL1500-E) quartz–halogen light source. The relative contribution of chlorophyll fluorescence quenching by photochemical and non-photochemical processes was estimated from the parameters q_P (ref. 29) and NPQ (ref. 30), respectively, after

illumination of the samples with continuous actinic light adjusted to approximate their growth irradiance.

Received 4 June; accepted 19 August 2004; doi:10.1038/nature02954.

- Dugdale, R. C. & Wilkerson, F. P. Silicate regulation of new production in the equatorial Pacific upwelling. *Nature* **391**, 270–273 (1998).
- Sunda, W. G., Swift, D. G. & Huntsman, S. A. Low iron requirement for growth in oceanic phytoplankton. *Nature* **351**, 55–57 (1991).
- Sunda, W. G. & Huntsman, S. A. Iron uptake and growth limitation in oceanic and coastal phytoplankton. *Mar. Chem.* **50**, 189–206 (1995).
- Ryther, J. H. & Kramer, D. D. Relative iron requirements of some coastal and offshore plankton algae. *Ecology* **42**, 444–446 (1961).
- Brand, L. E., Sunda, W. G. & Guillard, R. R. L. Limitation of marine phytoplankton reproductive rates by zinc, manganese, and iron. *Limnol. Oceanogr.* **28**, 1182–1198 (1983).
- Maldonado, M. T. & Price, N. M. Influence of N substrate on Fe requirements of marine centric diatoms. *Mar. Ecol. Prog. Ser.* **141**, 161–172 (1996).
- Raven, J. A. Predictions of Mn and Fe use efficiencies of phototrophic growth as a function of light availability for growth and of C assimilation pathway. *New Phytol.* **116**, 1–18 (1990).
- Raven, J. A., Evans, M. C. W. & Korb, R. E. The role of trace metals in photosynthetic electron transport in O_2 -evolving organisms. *Photosynth. Res.* **60**, 111–149 (1999).
- Quigg, A. *et al.* The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. *Nature* **425**, 291–294 (2003).
- Sunda, W. G. & Huntsman, S. A. Interrelated influence of iron, light and cell size on marine phytoplankton growth. *Nature* **390**, 389–392 (1997).
- Greene, R. M., Geider, R. J. & Falkowski, P. G. Effect of iron limitation on photosynthesis in a marine diatom. *Limnol. Oceanogr.* **36**, 1772–1782 (1991).
- Glover, H. Effects of iron deficiency on *Isochrysis galbana* (Chrysophyceae) and *Phaeodactylum tricornutum* (Bacillariophyceae). *J. Phycol.* **13**, 208–212 (1977).
- Kolber, Z., Zehr, J. & Falkowski, P. G. Effects of growth irradiance and nitrogen limitation on photosynthetic energy conversion in photosystem II. *Plant Physiol.* **88**, 923–929 (1988).
- Behrenfeld, M. J., Prasil, O., Kolber, Z. S., Babin, M. & Falkowski, P. G. Compensatory changes in Photosystem II electron turnover rates protect photosynthesis from photoinhibition. *Photosynth. Res.* **58**, 259–268 (1998).
- Horton, P., Ruban, A. V. & Walters, R. G. Regulation of light harvesting in green plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 655–684 (1996).
- Munekage, Y. *et al.* Cytochrome *b₆f* mutation specifically affects thermal dissipation of absorbed light energy in *Arabidopsis*. *Plant J.* **28**, 351–359 (2001).
- Price, G. D. *et al.* Chloroplast cytochrome *b₆f* and ATP synthase complexes in tobacco—transformation with antisense RNA against nuclear-encoded transcripts for the Rieske FeS and ATP-Δ polypeptides. *Aust. J. Plant Physiol.* **22**, 285–297 (1995).
- Lavaud, J., Rousseau, B., van Gorkom, H. J. & Etienne, A. L. Influence of the diadinoxanthin pool size on photoprotection in the marine planktonic diatom *Phaeodactylum tricornutum*. *Plant Physiol.* **129**, 1398–1406 (2002).
- Kühlheim, C., Ågren, J. & Jansson, S. Rapid regulation of light harvesting and plant fitness in the field. *Science* **297**, 91–93 (2002).
- MacIntyre, H. L., Kana, T. M. & Geider, R. J. The effect of water motion on short-term rates of photosynthesis by marine phytoplankton. *Trends Plant Sci.* **5**, 12–17 (2000).
- Anbar, A. D. & Knoll, A. H. Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science* **297**, 1137–1142 (2002).
- Hudson, R. J. M., Covault, D. T. & Morel, F. M. M. Investigations of iron coordination and redox reactions in seawater using ^{59}Fe radiometry and ion-pair solvent extraction of amphiphilic iron complexes. *Mar. Chem.* **38**, 209–235 (1992).
- Falkowski, P. G., Owens, T. G., Ley, A. C. & Mauzerall, D. C. Effects of growth irradiance levels on the ratio of reaction centers in two species of marine phytoplankton. *Plant Physiol.* **68**, 969–973 (1981).
- Bendall, D. S., Davenport, H. E. & Hill, R. Cytochrome components in chloroplasts of the higher plants. *Methods Enzymol.* **23A**, 327–344 (1971).
- Wood, P. M. Interchangeable copper and iron proteins in algal photosynthesis: studies on plastocyanin and cytochrome *c-552* in *Chlamydomonas*. *Eur. J. Biochem.* **87**, 9–19 (1978).
- Kurusu, G., Zhang, H., Smith, J. L. & Cramer, W. A. Structure of the cytochrome *b₆f* complex of oxygenic photosynthesis: tuning the cavity. *Science* **302**, 1009–1014 (2003).
- Jassby, A. D. & Platt, T. Mathematical formulation of the relationship between photosynthesis and light for phytoplankton. *Limnol. Oceanogr.* **21**, 540–547 (1976).
- Ley, A. C. & Mauzerall, D. C. Absolute absorption cross-sections for photosystem II and the minimum quantum requirement for photosynthesis in *Chlorella vulgaris*. *Biochim. Biophys. Acta* **680**, 95–106 (1982).
- Schreiber, U. Detection of rapid induction kinetics with a new type of high frequency modulated chlorophyll fluorometer. *Photosynth. Res.* **9**, 261–272 (1986).
- Bilger, D. H. & Björkman, O. Role of the xanthophyll cycle in photoprotection elucidated by measurements of light-induced absorbance changes, fluorescence and photosynthesis in *Hedera canariensis*. *Photosynth. Res.* **25**, 173–185 (1990).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. M. McKay for performing the ferredoxin and flavodoxin immunoassays, and J. Eaton-Rye, K. Hunter, M. Maldonado, S. Sander, P. Tortell and P. Boyd for comments on the manuscript. This work was supported by the Natural Sciences and Engineering Research Council of Canada and the Marsden Fund of New Zealand.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.F.S. (roberts@alkali.otago.ac.nz).

Pleiotropy as a mechanism to stabilize cooperation

Kevin R. Foster^{1*}, Gad Shaulsky², Joan E. Strassmann¹, David C. Queller¹ & Chris R. L. Thompson^{2,3*}

¹Ecology and Evolution, Rice University, Houston, Texas 77005, USA

²Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

³School of Biological Sciences, University of Manchester, Manchester M13 9PT, UK

* These authors contributed equally to this work.

Most genes affect many traits^{1–4}. This phenomenon, known as pleiotropy, is a major constraint on evolution because adaptive change in one trait may be prevented because it would compromise other traits affected by the same genes^{2,4}. Here we show that pleiotropy can have an unexpected effect and benefit one of the most enigmatic of adaptations—cooperation. A spectacular act of cooperation occurs in the social amoeba *Dictyostelium discoideum*, in which some cells die to form a stalk that holds the other cells aloft as reproductive spores^{5,6}. We have identified a gene, *dimA*⁷, in *D. discoideum* that has two contrasting effects. It is required to receive the signalling molecule DIF-1 that causes differentiation into prestalk cells. Ignoring DIF-1 and not becoming prestalk should allow cells to cheat by avoiding the stalk. However, we find that in aggregations containing the wild-type cells, lack of the *dimA* gene results in exclusion from spores. This pleiotropic linkage of stalk and spore formation limits the potential for cheating in *D. discoideum* because defecting on prestalk cell production results in an even greater reduction in spores. We propose that the evolution of pleiotropic links between cheating and personal costs can stabilize cooperative adaptations.

Acts of cooperation such as stalk formation in *D. discoideum* are a challenge for evolutionary biologists because of the potential for disruptive cheaters^{8–12}. When starving, the normally solitary amoebae of *D. discoideum* aggregate together to form a migratory slug and then a fruiting body in which most cells become spores but around one-fifth die to form a supporting stalk^{5,6}. Genetically different clones of *D. discoideum* will aggregate together and form chimaeric fruiting bodies⁵ and chimaerism seems to be common in nature because multiple clones co-occur in tiny soil samples¹³. This suggests that a defector¹² that produces fewer stalk cells can gain a selfish advantage⁵ and raises the question of how stalk formation is maintained^{9,11}. Around half of prestalk cells in *D. discoideum*, known as the prestalk O cells, are induced to differentiate by the signalling molecule DIF-1 (ref. 14), that is released by the neighbouring prespore cells¹⁵ (Fig. 1a). A viable cheating strategy, therefore, would be to ignore DIF-1 and overproduce spore cells in chimaeras. This is supported by the observation that cells with lowered DIF sensitivity, resulting from growth with glucose, cheat and are over represented in the spores when mixed with cells grown without glucose^{16,17}.

We used a knockout mutant to examine the effects of ignoring DIF-1. *dimA* encodes a central component of the DIF response pathway. A mutant with this gene disrupted (*dimA*[–]) ignores DIF-1 and produces prespore cells in place of prestalk O cells⁷ (Fig. 1a). We examined the behaviour of *dimA*[–] cells as they aggregated with equal numbers of their parental wild-type strain (AX4). The *dimA*[–] cells co-aggregated normally with AX4: *dimA*[–] and AX4 cells entered aggregations in equal numbers, and there was no loss of *dimA*[–] from aggregations during the migratory slug stage (mean percentage of *dimA*[–] cells in *dimA*/AX4 chimaeras from two time points in two independent experiments = 48%; 1 petri plate per mixture in each experiment, cells counted *N* = 1,168; Chi-squared

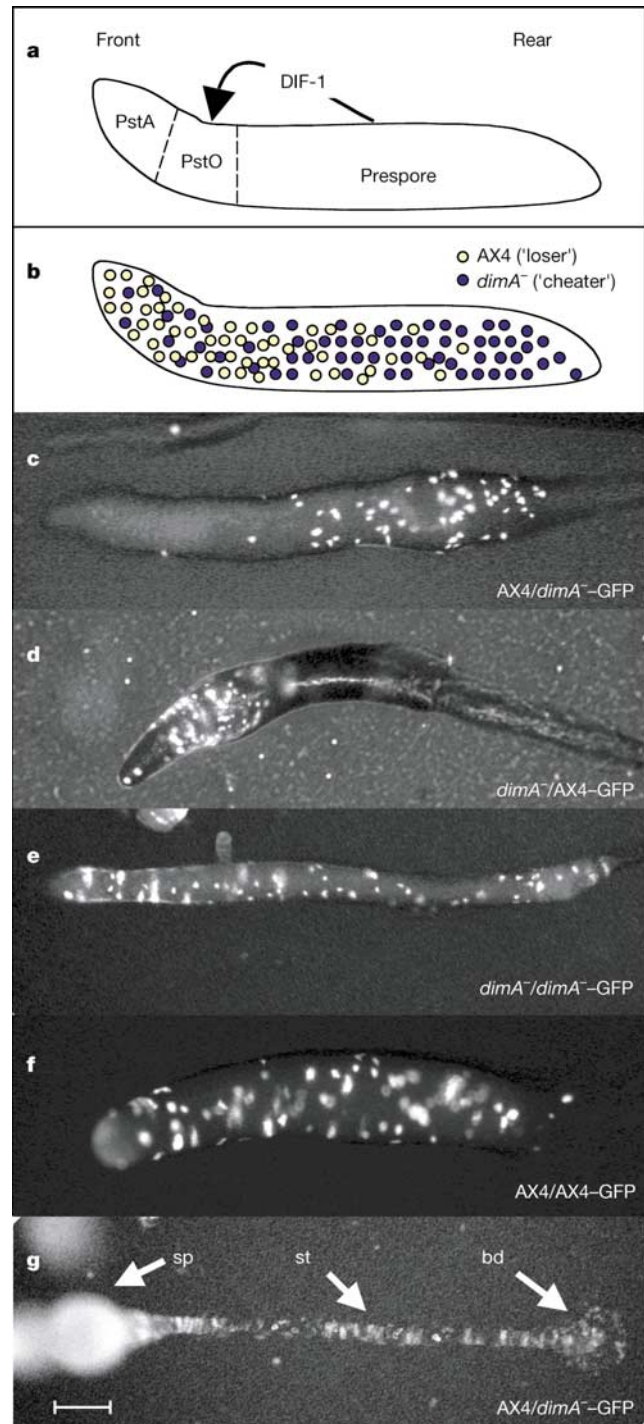


Figure 1 *dimA*[–] cells occupy the prespore zone and behave like a cheater in chimaeras. **a**, A schematic diagram of a slug of the slime mould *Dictyostelium discoideum* showing three of the major cell types: prestalk A (pstA), prestalk O (pstO) and prespore cells. Prespore cells release DIF-1, which induces pstO cell differentiation. **b**, Schematic diagram showing the distribution of *dimA*[–] cells in a chimaeric slug with the wild-type AX4. **c**, *dimA*[–] cells are over represented in the rear of the slug: chimaeras of 50% AX4, 48% *dimA*[–] and 2% *dimA*[–]–GFP. Some *dimA*[–] cells are present in the prestalk zone (see text) but are not visible with 2% GFP expressing cells. **d**, AX4 cells are over represented in the front of the slug: 50% *dimA*[–], 48% AX4 and 2% AX4–GFP. **e**, Control: 98% *dimA*[–] and 2% *dimA*[–]–GFP. **f**, Control: 98% AX4 and 2% AX4–GFP. **g**, A chimaeric fruiting body (50% AX4, 48% *dimA*[–] and 2% *dimA*[–]–GFP): sp, sorus; st, stalk; bd, basal disc. There are more *dimA*[–] cells in the stalk than basal disc. Scale bar is 0.1 mm.

from 50% = 2.23; $P = 0.13$). As predicted, $dimA^-$ defected by preferentially sorting to the prespore zone (Fig. 1b–f). Cell counts of the prestalk region of slugs (front 25%) confirmed there were significantly fewer $dimA^-$ cells than AX4 cells (mean proportion $dimA^-$ in chimaeras with AX4 cells in two experiments = 34%; 10 slugs per mixture in each experiment; cells counted $N = 1,239$; Chi-squared versus 50% = 126.93; $P < 0.0001$). The localization of $dimA^-$ cells at the back of slugs is not because they are weakened cells with low motility. Low motility cells would be expected to be preferentially sloughed during slug migration. In contrast, we found equal numbers of $dimA^-$ and AX4 cells in slugs that had migrated (above) and did not observe an excess of $dimA^-$ cells being left behind in slug trails. In addition, $dimA^-$ cells in the back of the slug are true prespore cells because they express the prespore marker *cotB-lacZ*⁷. Finally, $dimA^-$ cells did not seem to be simply impaired in their ability to differentiate. Mutants which exhibit general differentiation defects are typically greatly enriched in the basal disc of chimaeric fruiting bodies with wild-type cells¹⁸. In contrast, $dimA^-$ cells contributed to all terminally differentiated cell types and were not enriched in the basal disc of chimaeras (Fig. 1g).

To test whether the defection of $dimA^-$ in slugs resulted in a competitive advantage, we examined the spores in the fruiting bodies that form after the slug stage. We were surprised to find that $dimA^-$ cells ultimately lose out with about half as many present as AX4 cells in the spores (percentage of green fluorescent protein (GFP) spores in AX4/ $dimA^-$ -GFP chimaeras = 34.7%; spores counted = 1,256; Student's t -test versus 50%: $t = -6.4$; $N = 6$ petri plates; $P < 0.002$; percentage GFP spores in $dimA^-$ /AX4-GFP chimaeras = 63%; spores counted = 1,279, Student's t -test versus 50%: $t = 4.2$; $N = 6$ petri plates; $P < 0.009$; Fig. 2). In contrast, controls of $dimA^-$ / $dimA^-$ -GFP chimaeras and AX4/AX4-GFP chimaeras demonstrated no significant biases from 50:50 owing to GFP labelling (percentage of GFP spores from both controls = 52%; spores counted = 1,275; Student's t -test versus 50%: $t = 1.164$; $N = 12$ petri plates; $P = 0.27$). The low number of $dimA^-$ spores is not simply because $dimA^-$ cells present in the spore head fail to make spores and remain as amoebae: there was no reduction in total spore number in the $dimA^-$ /AX4 chimaera compared with AX4 alone (Fig. 3). Additional evidence that $dimA^-$ losing is not because of an intrinsic sporulation deficit comes from clonal $dimA^-$ development, which shows that $dimA^-$ is effective at producing spores: $dimA^-$ produces around 82% of the spore number of the wild type (mean spores per plate: $dimA^- = 1.70 \times 10^7$; AX4 = 2.08×10^7 ; $N = 36$ plates). Taken together, these results suggest that in chimaeras $dimA^-$ cells are

replaced in the spore population by AX4 late in development.

To test whether AX4 prestalk cells replace $dimA^-$ prespore cells, we used a prestalk specific marker *ecmA-lacZ*¹⁹. This marker is expressed in about one-fifth of prestalk cells (data not shown) but is rarely expressed in spore cells. However, prestalk cells that express the gene and then transdifferentiate into spores will retain expression owing to the inherent stability of the β -galactosidase enzyme produced by the *lacZ* gene²⁰. The number of *lacZ* positive AX4 spores increased sixfold when developed in chimaera with $dimA^-$ cells (Fig. 4). This result is consistent with the majority of AX4 prestalk cells switching fate and becoming spores late in development and replacing $dimA^-$ cells in the spores (for example, if 12.5% of cells are AX4 prestalk and one-fifth of these express *lacZ* and all transdifferentiate, we expect about 3% of cells to be labelled spores, not far above the percentage shown in Fig. 4).

The *dimA* gene therefore has two contrasting effects. Cells that do not express *dimA* behave like a cheater in chimaeric slugs by increasing their representation in the prespore population, presumably because they are able to ignore the signal of prestalk cell induction, DIF-1. In the terminology of Velicer¹², $dimA^-$ cells are defectors that under-invest in the cooperative resource of prestalk cells. However, they do not benefit from defection because of a second phenotype of $dimA^-$ cells that causes them to be competitively excluded from the spores late in development. This second pleiotropic effect of *dimA* makes ignoring the DIF-1 signal unprofitable and helps to limit cheating and ensure fair contribution to the stalk.

We do not expect all cheating to be controlled by pleiotropy and single-gene mutants have been found that cheat under laboratory conditions¹². Nevertheless, we predict that pleiotropy will be a common form of cheater control because it is so ubiquitous. It has been neglected because the organisms studied by most sociobiologists are genetically intractable, but additional evidence can be found from microorganisms. For example, another potential cheating strategy in *D. discoideum* is to reduce adhesion and slide to the back of the slug, where the prespore zone is located (Fig. 1a). This is supported by the *csA*⁻ knockout mutant, which lacks a key adhesion molecule and is a cheater in chimaeras with the wild type on agar^{21,22}. However, its reduced adhesion means that it has trouble getting into aggregations on soil and incurs a great fitness cost²¹. Another example comes from the fluorescent bacterium *Vibrio fischeri*, that is used for bio-illumination by the squid *Euprymna scolopes*. Bacteria that defect by not producing the light production enzyme luciferase should gain a selfish advantage. However, a luciferase knockout strain grows poorly in the squid light organ²³ suggesting a pleiotropic link between light production

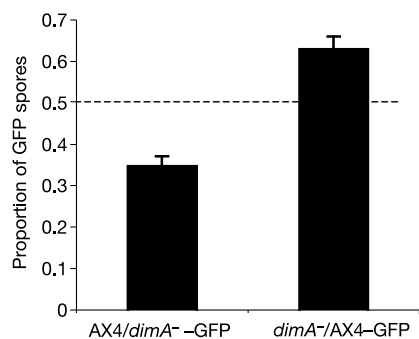


Figure 2 The $dimA^-$ mutant loses out in the fruiting body. In the spore head of $dimA^-$ /AX4 chimaeras, there are almost twice as many AX4 spores as $dimA^-$ spores. Data are from two independent experiments. Bars are standard errors calculated using each petri plate as a single replicate. In addition, we have independently confirmed the losing behaviour of $dimA^-$ without GFP expression, using blasticidin S resistance (C.R.T. and L. Santorelli, unpublished data) to distinguish between AX4 and $dimA^-$ cells⁷.

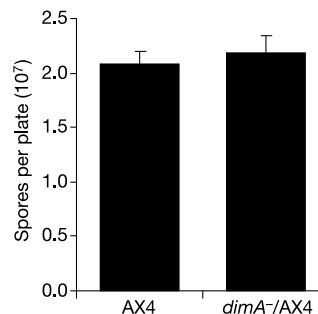


Figure 3 AX4/ $dimA^-$ chimaeras produce the same number of spores as wild type (AX4) alone. Total number of spores produced per plate by AX4 clones and chimaeras of AX4/ $dimA^-$ (AX4 = 2.08×10^7 spores, $dimA^-$ /AX4 chimaeras = 2.18×10^7 , $N = 30$ petri plates, t -test = -0.523 , $P = 0.61$). Bars are standard errors calculated using each petri plate as a single replicate. In addition, spore viability assays demonstrated no significant difference between AX4 and $dimA^-$ /AX4 chimaeras (t -test, $t = 1.264$, $N = 30$ petri plates, $P = 0.22$).

and growth. Finally, in *Escherichia coli* bacteria, a single-gene GASP (growth advantage under stationary phase) mutant has been identified that defects from the coordinated developmental process that naturally occurs under low nutrients²⁴. Instead of reducing reproduction to form resistant spore-like cells, it rapidly divides and replaces the wild type. However, defection again results in a cost because the resulting GASP cells do not tolerate acidity and cannot survive passage through their mammalian hosts' stomachs.

A form of cheater control related to pleiotropy occurs by tight linkage of genes in colicin production in bacteria²⁵. Colicins are produced by bacteria through a plasmid that contains genes for both colicin production and resistance to their toxic effects. This linkage of genes prevents loss of colicin production by plasmid loss because this would also cause loss of resistance²⁶. However, it is unknown if colicin genes exhibit pleiotropic constraints and loss of production may occur without loss of resistance through mutation of the colicin production gene alone²⁵.

Pleiotropy may be shown as important in more familiar social organisms such as the social insects, once molecular mechanisms are uncovered. In several species, the queen releases a pheromone that suppresses worker ovary development, suggesting chemical control of reproduction²⁷. The control hypothesis has been criticized on theoretical grounds because workers should evolve to ignore any chemical control agent²⁷. However, workers could be prevented from ignoring the queen pheromone if, analogous to DIF-1, it targets a pleiotropic gene that is required for both response to the queen pheromone and ovary development itself.

The spread of social genes will be promoted by mechanisms that make their loss costly ('intrinsic defector inferiority')²⁶. Our study demonstrates that pleiotropy is one such mechanism. The complex interrelationships of developmental and biochemical processes mean that genes have innumerable idiosyncratic side effects that constrain adaptation¹⁻⁴. Paradoxically, we have found that these constraints can benefit cooperative adaptations by limiting the potential for individual selfishness. Social innovations will persist only when disruptive cheaters cannot invade^{28,29}. As a result, the successful innovations will often be tied through pleiotropy to other essential functions, limiting the number of cheating mutations with a net advantage. Pleiotropy provides one way to limit individual rebellion and allow stable cooperation to evolve. Where pleiotropic costs are strong, other means of control like high relatedness³⁰ and policing¹¹ may be less necessary. □

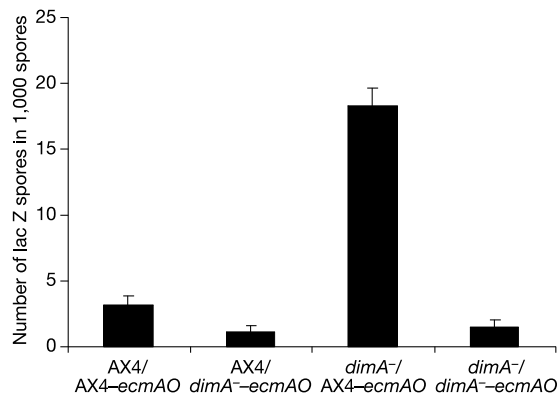


Figure 4 The majority of AX4 prestalk cells switch fate and become prespore cells when in chimaeras with *dimA*⁻. There is a sixfold increase in the number of spores expressing the prestalk marker *ecmAO-lacZ* in chimaeras of *dimA*⁻/AX4-*ecmAO-lacZ* compared to other mixtures (analysis of variance between groups (ANOVA), $N = 24$, $F = 70.9$, $P < 0.001$). Data are from two counts of 1,000 spores from each of three independent experiments. Bars are standard errors calculated using each count as a single replicate.

Methods

Cell growth, maintenance and chimeric development

Strains were maintained in association with *Klebsiella aerogenes* or in liquid medium⁷. *Actin15-GFP* and *ecmAO-lacZ* transformants were generated as described⁷ with initial selection in 20 µg G418 and maintenance in 2.5 µg G418. Before development, cells were grown for 48 h without G418. For development, cells were collected by centrifugation, washed with KK2 (16.1 mM KH₂PO₄ and 3.7 mM K₂HPO₄) and developed at a density of 6.4×10^5 cells per cm² on KK2 plates containing 1.5% purified agar (Oxoid)⁷.

Disaggregation, spore counts and lacZ staining of spores

GFP labelling was used to distinguish between *dimA*⁻ and AX4 cells. To count cells in chimaeras during development, multicellular structures were disaggregated at two time points: mounds and migrated slugs. The structures were washed from plates with KK2 and separated from loose cells using a 77 µm cell sieve. The structures were then transferred from the sieve into 20 mM EDTA and passed several times through a 20-gauge needle in order to disaggregate the cells for counting by fluorescence microscopy. For counting of spores, structures were harvested from plates after 48 h of development into 20 mM EDTA containing 0.1% NP-40 to lyse amoebae. Spore density was counted using a haemocytometer. Fluorescence microscopy was used to count GFP spores (Fig. 2), and pure *dimA*⁻-GFP and pure AX4-GFP controls used to adjust counts for non-glowing cells. For the *lacZ* labelled strains, slugs of pure strains of AX4-*ecmAO-lacZ* and *dimA*⁻-*ecmAO-lacZ* were disaggregated as described above and stained to give percentage of expressing cells. Spore staining was performed as described²⁰. To measure spore viability, the colonies that resulted from the plating of 100 spores were counted. All experiments were performed at least twice and data are combined because there were no significant differences between experiments.

Received 23 April; accepted 30 July 2004; doi:10.1038/nature02894.

- Wright, S. *Evolution and the Genetics of Populations* Vol. 1 (Univ. of Chicago Press, Chicago, 1968).
- Barton, N. H. Pleiotropic models of quantitative variation. *Genetics* **124**, 773–782 (1990).
- Featherstone, D. E. & Brodie, K. Wrestling with pleiotropy: genomic and topological analysis of the yeast expression network. *BioEssays* **24**, 267–274 (2002).
- Otto, S. P. Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc. R. Soc. Lond. B* **271**, 705–714 (2004).
- Strassmann, J. E., Zhu, Y. & Queller, D. C. Altruism and social cheating in the social amoeba *Dictyostelium discoideum*. *Nature* **408**, 965–967 (2000).
- Kessin, R. H. *Dictyostelium: Evolution, Cell Biology and the Development of Multicellularity* (Cambridge Univ. Press, Cambridge, 2001).
- Thompson, C. R. L., Fu, Q., Buhay, C., Kay, R. R. & Shaulsky, G. A bZIP/bRLZ transcription factor required for DIF signaling in *Dictyostelium*. *Development* **131**, 513–523 (2004).
- Maynard Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (Freeman, New York, 1995).
- Atzmöy, D., Zahavi, A. & Nanjundiah, V. Altruistic behaviour in *Dictyostelium discoideum* explained on the basis of individual selection. *Curr. Sci.* **72**, 142–145 (1997).
- Keller, L. (ed.) *Levels of Selection in Evolution* (Princeton Univ. Press, Princeton, New Jersey, 1999).
- Frank, S. A. Repression of competition and the evolution of cooperation. *Evolution* **57**, 693–705 (2003).
- Velicer, G. J. Social strife in the microbial world. *Trends Microbiol.* **11**, 330–337 (2003).
- Fortunato, A., Strassmann, J. E., Santorelli, L. & Queller, D. C. Co-occurrence in nature of different clones of the social amoeba, *Dictyostelium discoideum*. *Mol. Ecol.* **12**, 1031–1038 (2003).
- Thompson, C. R. L. & Kay, R. R. The role of DIF-1 signaling in *Dictyostelium* development. *Mol. Cell* **6**, 1509–1514 (2000).
- Kay, R. R. & Thompson, C. R. L. Cross-induction of cell types in *Dictyostelium*: evidence that DIF-1 is made by prespore cells. *Development* **128**, 4959–4966 (2001).
- Leach, C. K., Ashworth, J. M. & Garrod, D. R. Cell sorting out during the differentiation of mixtures of metabolically distinct populations of *Dictyostelium discoideum*. *J. Embryol. Exp. Morphol.* **29**, 647–661 (1973).
- Thompson, C. R. L. & Kay, R. R. Cell-fate choice in *Dictyostelium*: intrinsic biases modulate sensitivity to DIF signaling. *Dev. Biol.* **227**, 56–64 (2000).
- Insall, R. et al. CRAC, a cytosolic protein containing a pleckstrin homology domain, is required for receptor and G protein-mediated activation of adenylyl cyclase in *Dictyostelium*. *J. Cell Biol.* **126**, 1537–1545 (1994).
- Jermyn, K. A. & Williams, J. G. An analysis of culmination in *Dictyostelium* using prestalk and stalk-specific cell autonomous markers. *Development* **111**, 779–787 (1991).
- Shaulsky, G. & Loomis, W. F. Cell type regulation in response to expression of ricin A in *Dictyostelium*. *Dev. Biol.* **160**, 85–98 (1993).
- Ponte, E., Bracco, E., Faix, J. & Bozzaro, S. Detection of subtle phenotypes: the case of the cell adhesion molecule csA in *Dictyostelium*. *Proc. Natl Acad. Sci. USA* **95**, 9360–9365 (1998).
- Queller, D. C., Ponte, E., Bozzaro, S. & Strassmann, J. E. Single-gene greenbeard effects in the social amoeba, *Dictyostelium discoideum*. *Science* **299**, 105–106 (2003).
- Visick, K. L., Foster, J., Doi, J., McFall-Ngai, M. & Ruby, E. G. *Vibrio fischeri lux* genes play an important role in colonization and development of the host light organ. *J. Bacteriol.* **182**, 4578–4586 (2000).
- Vulic, M. & Kolter, R. Evolutionary cheating in *Escherichia coli* stationary phase cultures. *Genetics* **158**, 519–526 (2001).
- Riley, M. A. & Wertz, J. E. Bacteriocins: evolution, ecology, and application. *Annu. Rev. Microbiol.* **56**, 117–137 (2002).
- Travisano, M. & Velicer, G. J. Strategies of microbial cheater control. *Trends Microbiol.* **12**, 72–78 (2004).
- Keller, L. & Nonacs, P. The role of queen pheromones in social insects: queen control or queen signal? *Anim. Behav.* **45**, 787–794 (1993).
- Hurst, L. D., Atlan, A. & Bengtsson, B. O. Genetic conflicts. *Q. Rev. Biol.* **71**, 317–364 (1996).
- Nunney, L. in *Levels of Selection in Evolution* (ed. Keller, L.) 238–252 (Princeton Univ. Press, Princeton, New Jersey, 1999).

30. Hamilton, W. D. The genetical evolution of social behavior I and II. *J. Theor. Biol.* 7, 1–52 (1964).

Acknowledgements Thank you to N. Mehdiabadi and G. Velicer for helpful comments. This material is based upon work supported by the National Science Foundation, and C.R.L.T. was supported by a Wellcome Trust International Prize Travelling Research Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.R.F. (krfooster@rice.edu) or C.R.L.T. (christopher.thompson@man.ac.uk).

Migratory neural crest-like cells form body pigmentation in a urochordate embryo

William R. Jeffery, Allen G. Strickler & Yoshiyuki Yamamoto*

Department of Biology, University of Maryland, College Park, Maryland 20742, USA

* Present address: Evolutionary Anatomy Unit, Department of Anatomy & Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK

The neural crest, a source of many different cell types in vertebrate embryos, has not been identified in other chordates^{1–3}. Current opinion therefore holds that neural crest cells were a vertebrate innovation^{4–7}. Here we describe a migratory cell population resembling neural crest cells in the ascidian urochordate *Ecteinascidia turbinata*. Labelling of embryos and larvae with the vital lipophilic dye DiI enabled us to detect cells that emerge from the neural tube, migrate into the body wall and siphon primordia, and subsequently differentiate as pigment cells. These cells express HNK-1 antigen and *Zic* gene markers of vertebrate neural crest cells. The results suggest that migratory cells with some of the features of neural crest cells are present in the urochordates. Thus, we propose a hypothesis for neural crest evolution beginning with the release of migratory cells from the CNS to produce body pigmentation in the common ancestor of the urochordates and vertebrates. These cells may have gained additional functions or were joined by other cell types to generate the variety of derivatives typical of the vertebrate neural crest.

Neural crest cells delaminate from the dorsal neural tube in an anterior to posterior sequence, migrate through stereotypical pathways, and differentiate into a variety of cell types in vertebrate embryos, including neuronal, glial, endocrine, skeletal and pigment cells^{1,2}. Migratory neural crest cells have not been described in non-vertebrate chordates. In ascidian urochordates and amphioxus, however, cells at the border of the neural plate or within the neural tube are often proposed as evolutionary precursors of the neural crest^{3,7–10}.

The life cycle of the ascidians is usually separated into distinct larval (or embryonic) and adult developmental phases¹¹. The tadpole larva consists of a trunk (or head) with a dorsal central nervous system (CNS) containing two melanized sensory organs (otolith and ocellus) and a tail with a notochord, spinal cord and muscle cells. During metamorphosis, the larval tail is destroyed and the head is extensively reorganized into a sessile filter-feeding adult. Previous searches for neural crest cells were restricted to the embryonic phase in ascidian species such as *Ciona intestinalis*, whose small larvae exhibit the conventional mode of development. These studies did not use cell-tracing methods, which originally defined the neural crest in vertebrates^{1,2}. We have investigated the possibility of neural crest-like cells in urochordates using the

colonial ascidian *Ecteinascidia turbinata* (Fig. 1a), which exhibits a giant tadpole that is suitable for cell-tracing approaches to detect migratory cells (Fig. 1b). In contrast to more commonly studied ascidians, the *Ecteinascidia* tadpole initiates adult development in its head during the embryonic phase, where adult features such as the pharynx, heart, siphons and stellate body pigment cells are formed precociously (Fig. 1c, d)^{12,13}.

Using the otolith and ocellus as landmarks, DiI was injected into the anterior neural tube at the early tailbud stage and the injected embryos were subsequently examined by fluorescence microscopy. There was no movement of DiI or labelled cells immediately after injection (Fig. 2a, b). By 4 h after injection, however, DiI-labelled cells were observed to migrate from the injection site towards the developing siphons (Fig. 2c, d). Migratory cells continued to be observed through the late tailbud stage (Fig. 2e, f). Some of these migrating cells had the stellate morphology typical of body pigment cells (Fig. 2f). At the mid-tailbud stage, DiI was injected into the anterior neural tube, the posterior neural tube, or the ventral head epidermis. As before, migrating cells emerged from the anterior neural tube and moved ventrolaterally into the body wall (Fig. 2g). Sectioning showed that cell migration occurred in two pathways: (1) through the dorsal mesoderm surrounding the neural tube, and (2) between the mesoderm and epidermis (Fig. 2j–l). In contrast, DiI-labelled cells did not migrate from the posterior neural tube (Fig. 2h) or the ventral epidermis (Fig. 2i), although local rearrangements of labelled cells occurred in these regions. The results show that migratory cells emerge from the neural tube during *E. turbinata* embryogenesis.

DiI-labelling experiments were continued during later development. After DiI injection into the posterior neural tube at the late-tailbud stage, labelled cells migrated into the head but were excluded from the tail (Fig. 3a, b). At the same stage, migratory cells were still generated after DiI injection into the anterior neural tube, and many of these cells became associated with the developing siphons (see Fig. 2e, f; data not shown). These results suggest that migratory cells emerge in an anterior to posterior sequence during *Ecteinascidia* development. In contrast to vertebrate neural crest cells, however, the

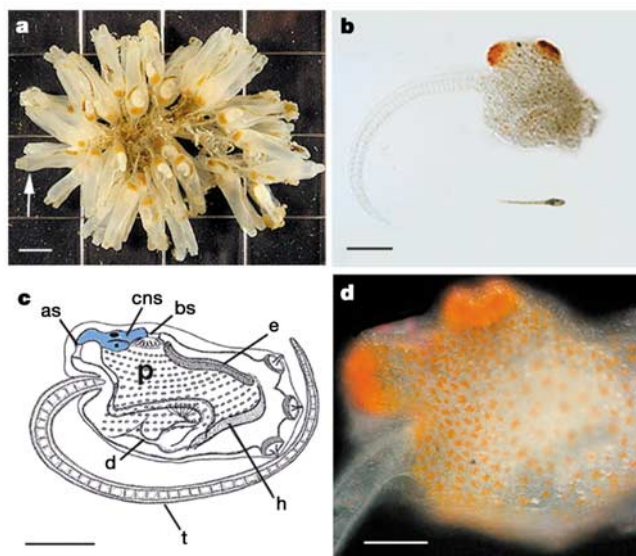


Figure 1 *Ecteinascidia turbinata*. **a**, A colony showing gravid zooids (arrow). **b**, The giant *Ecteinascidia* tadpole (above) compared to a small *Styela clava* tadpole (below). **c**, A diagram of the *Ecteinascidia* tadpole after ref. 12, showing the CNS (blue) with melanized sensory cells (black), branchial siphon (bs), atrial siphon (as), endostyle (e), heart (h), perforated pharynx (p), digestive tract (d), and tail (t). **d**, Orange pigment cells distributed throughout the body wall and concentrated in the siphon primordia of an *Ecteinascidia* tadpole. Scale bars: **a**, 1 cm; **b–d**, 800 μ m.

Ecteinascidia cells migrated singly rather than in streams (Figs 2, 3), possibly owing to the absence of segmentation in the ascidian trunk. Also in contrast to vertebrate neural crest cells, the ascidian migratory cells appeared to emerge from all dorsal–ventral levels of the neural tube/CNS and for an extended time period after its formation from the neural plate. Injection of DiI into the CNS of mature tadpoles and examination of post-metamorphic adults by fluorescence microscopy detected labelled cells throughout the body wall and siphons (Fig. 3c–f). Most of the labelled cells were identified as pigment cells by their orange colour and stellate morphology (Figs 2f and 3g, h): in a single animal 28 of 36 DiI-labelled cells coincided exactly with a pigment cell. The remaining labelled cells were located near a pigment cell but their identity could not be determined with certainty. The results suggest that the migratory cells differentiate primarily into body pigment cells.

To determine whether the migratory cells express neural crest markers, embryos were stained with HNK-1 antibody or subjected to *in situ* hybridization with a *Zic* (*EtZic*) gene probe. HNK-1 antibody recognizes a surface glycolipid that is expressed in vertebrate neural crest cells during the migration phase but disappears shortly after terminal differentiation^{14–17}. In *Ecteinascidia* larvae, HNK-1 expression was detected in dorsal body wall and siphon cells showing the typical morphology of pigment cells (Fig. 4a–d, h). *Zic* family genes encode C2H2 zinc finger transcription factors that are expressed in the neuroectoderm and migrating neural crest cells of vertebrate embryos and can lead to ectopic neural crest formation

when misexpressed during *Xenopus* development^{18–20}. As previously described in *Ciona* and amphioxus embryos^{21,22}, *Zic* was expressed throughout the *Ecteinascidia* neural tube at early developmental stages (data not shown). By the late-tailbud stage, however, *EtZic* expression was confined to dorsal body wall cells with typical pigment cell morphology (Fig. 4e, g). Cells expressing HNK-1 and *EtZic* were abundant in the developing siphons, which also accumulate migratory cells and have been suggested to be homologous to vertebrate placodes^{23,24}. The results demonstrate the existence of migratory cells with several striking similarities to neural crest cells in a urochordate embryo. These cells emerge from the neural tube/CNS in an anterior to posterior sequence, migrate below the epidermis and through the mesoderm, express HNK-1 and *Zic* markers, differentiate into somatic pigment cells, and contribute to placode-like structures.

All vertebrate body pigment cells, including melanophores, iridophores, and xanthophores (which are most similar to *Ecteinascidia* orange pigment cells), are derived from the neural crest²⁵. Although most DiI-labelled cells appear to differentiate into pigment cells, the possibility that some of these cells may have additional fates cannot be excluded. The migratory pigment cells we have identified are restricted to the developing head of *Ecteinascidia* larvae, suggesting that they may be special features of the adult rather than the larval development phase. This interpretation is consistent with evidence for smaller-scale migration of adult neuronal precursors from the CNS during ascidian meta-

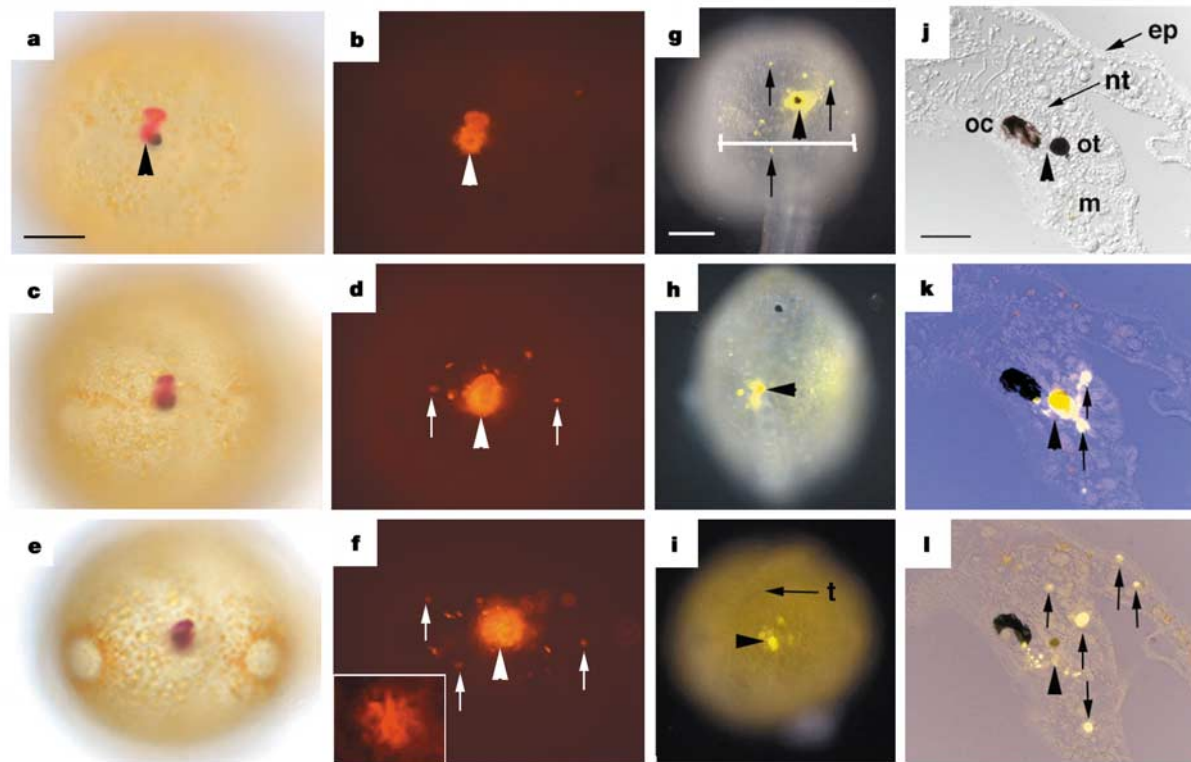


Figure 2 DiI tracing of migratory cells in embryos. **a–f**, Bright-field (**a, c, e**) and fluorescence (**b, d, f**) images of the same embryo injected in the anterior neural tube near the otolith/ocellus (black dot) at the early-tailbud stage. Inset in **f**, High magnification of stellate migrating cell. Images shown at 10 min (**a, b**), 4 h (**c, d**), and 14 h (**e, f**) after injection. **g–i**, Fluorescence images of DiI-labelled cells after injection into the anterior neural tube near the otolith/ocellus (**g**), the posterior neural tube region near the junction of the trunk and tail (**h**), and the ventral epidermis (**i**) at the mid-tailbud stage. The white bracketed horizontal line in **g** indicates the plane of section in **j–l**. **j–l**, Bright-field (**j**) and

fluorescence (**k, l**) images of a mid-tailbud embryo cross-sectioned serially with respect to the anterior–posterior axis showing the injection site between the otolith (ot) and ocellus (oc) and labelled cells migrating from the neural tube (nt) into the dorsal mesoderm (m) and below the epidermis (ep) (arrows). Migratory cells were observed leaving the injection site in every case ($n = 14$) in which DiI was injected into the anterior neural tube. Views are dorsal in **a–h** and ventral in **i**; **a–f**, anterior on the right; **g, h**, anterior on top. Arrowheads, DiI injection sites. Arrows, DiI-labelled migrating cells. Scale bars: **a, g**, 200 μm ; **j**, 30 μm ; magnification is the same in **a–f**, **g–i** and **j–l**.

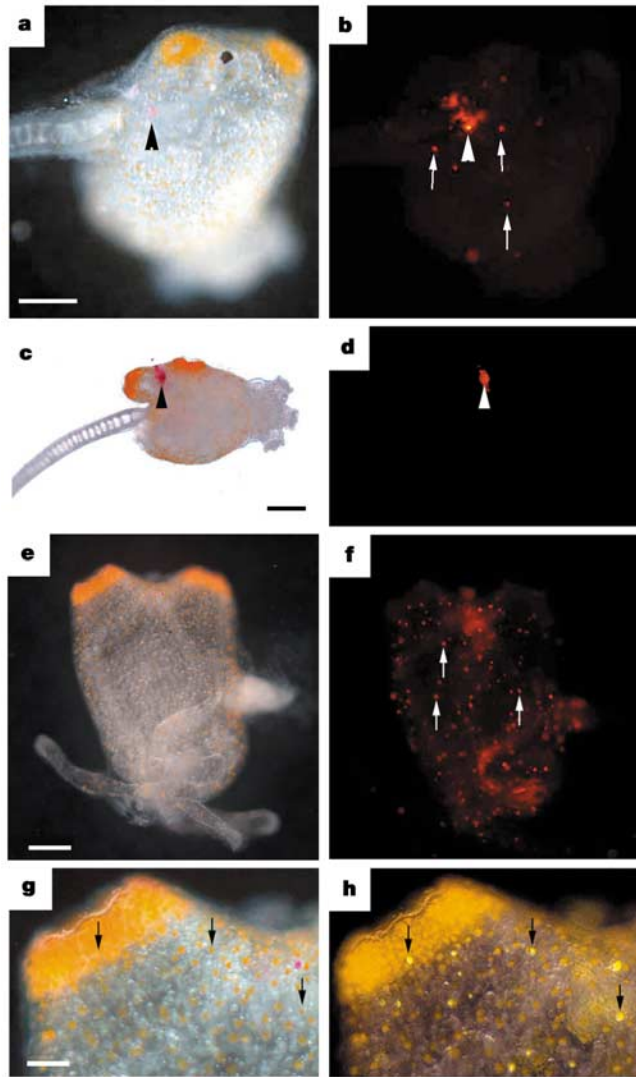


Figure 3 Dil tracing and fate of migratory cells in developing tadpoles and post-metamorphic adults. **a, b**, Bright-field (**a**) and fluorescence (**b**) images of the same embryo injected with Dil in the posterior neural region at the late-tailbud stage. **c–f**, Bright-field (**c, e**) and fluorescence (**d, f**) image of a Dil-injected mature tadpole (**c, d**) showing labelled cells 12 h later after metamorphosis (**e, f**). **g, h**, Fate of Dil-labelled cells in a post-metamorphic adult. **g**, Bright-field image showing individual orange pigment cells in the body wall. **h**, Fluorescence image of the same animal as in **g** showing Dil-labelled cells (in yellow). All views are lateral. Arrows indicate some of the orange pigment cells that also show Dil labelling. Scale bars: **a**, 250 μm ; **c**, 200 μm ; **e**, 100 μm ; **g**, 20 μm ; magnification is the same in **a** and **b**, **c** and **d**, **e** and **f**, and **g** and **h**.

morphosis^{26–28} and the presence of calcitonin-producing cells, a vertebrate neural crest derivative^{1,2}, in the adult pharynx²⁹.

We conclude that migratory neural crest-like cells are present in the ascidian *Ecteinascidia*. Although these cells could have evolved independently within the urochordate line or even within *Ecteinascidia*, HNK-1 and *Zic* expression suggests that they are likely to share a common origin with vertebrate neural crest cells. In support of a common evolutionary heritage of ascidian migratory cells and vertebrate neural crest cells, one of us (W.R.J.; manuscript in preparation) has recently discovered similar HNK-1 positive cells in four other ascidian species from evolutionarily diverse families. We therefore propose the following scenario for neural crest evolution in the chordates.

First, the neural tube/CNS attained the capacity to generate

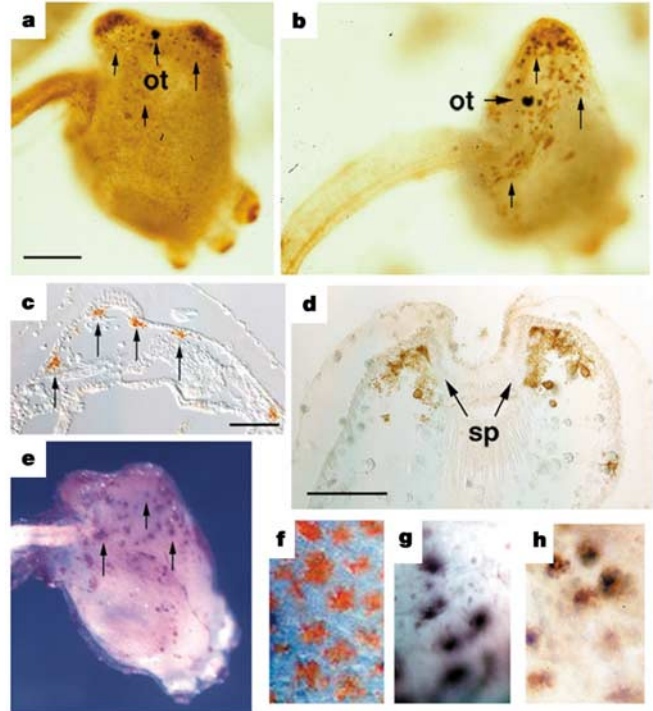


Figure 4 Neural crest markers. Expression of HNK-1 antigen (**a–d, h**) and the *EtZic* gene (**e, g**) in developing larvae. **a, b**, Lateral (**a**) and dorsal (**b**) views showing HNK-1 expressing cells (arrows). **c, d**, Sections through the dorsal body wall (**c**) and oral siphon primordium (sp) (**d**) showing HNK-1-stained cells (arrows). **e**, Lateral view of *in situ* hybridized larva showing *EtZic*-expressing cells (arrows) in the dorsal body wall. **f–h**, Comparison of orange pigment cells (**f**), *EtZic*-expressing cells (**g**), and HNK-1-stained cells (**h**) in the larval body wall. Scale bars: **a**, 250 μm ; **c, d**, 30 μm ; magnification is the same in **a, b** and **e**; **f–h** are $\times 10$ the magnification of **a**.

migratory pigment cells, possibly as a means to protect a sessile ascidian-like chordate ancestor from the harmful effects of sunlight in shallow marine habitats. These migratory pigment-generating cells may have been lost secondarily in amphioxus, which live buried in marine sediment and lack body pigmentation. Alternatively, it is conceivable that urochordates are the true sister group of vertebrates and that the migratory cells evolved in their common ancestor after the divergence of amphioxus from the chordate lineage. Later in chordate evolution, probably at or near the base of the vertebrate radiation, the primitive migratory cells gained additional functions and/or were joined by other cell types to generate the multiple derivatives characteristic of the neural crest. □

Methods

Animals

Ecteinascidia turbinata was obtained from Gulf Specimens, Inc. or collected near the Bermuda Biological Station, St Georges West, Bermuda. Tailbud-stage embryos and larvae were dissected from viviparous adults and cultured in Millipore-filtered sea water (MFSW) at 19 °C. The chorion of pre-hatching stage embryos was removed by dissection with sharpened tungsten needles.

Dil cell tracing

Dechorionated embryos and hatched larvae were immobilized for microinjection in 0.2% agar/MFSW. The vital lipophilic dye 1, 1'-didodecyl-3, 3, 3'-tetramethylindocarbocyanine perchlorate (DiI) (C-7000, Cell tracker, Molecular Probes) was dissolved to a final concentration of 37.5 $\mu\text{g ml}^{-1}$ in 92% *N, N*-dimethylformamide/8% 2 $\times$ distilled water by briefly heating to 50 °C, and glass micropipettes were back-filled with the DiI solution. Before DiI injection a small puncture was made with a tungsten needle in the larval tunic above the preferred DiI injection site. The DiI-containing micropipette was then inserted into the puncture and a bolus of DiI was pressure-injected. After microinjection, the embryos and larvae were washed several times in MFSW, and viewed with bright-field and

fluorescence optics. The DiI-injected embryos were fixed in 4% paraformaldehyde, embedded in polyester wax, sectioned at 10 μ m, and sections were viewed unstained with bright-field and fluorescence optics.

HNK-1 staining

Larvae were fixed in 4% paraformaldehyde in PBS for 30 min at room temperature and then in 100% methanol for 5 min at -20°C . The fixed larvae were washed in PBS containing 1% Triton X-100, and stained with a 1:20 dilution of HNK-1 antibody (Biopharmagen) for 24 h at 4°C using the Vectastain ABC Peroxidase kit (Vector Laboratories). Antibody incubations and blocking were carried out in 2% Superblock and controls used non-immune mouse serum.

EtZic isolation and *in situ* hybridization

The oligodeoxynucleotide primers AARGCMAATACAARCTATCAACC (forward) and ACYTTCATGTGCTTSCCKRAGRGAGC (reverse) were used to isolate part of an *Ecteinascidia* Zic gene by polymerase chain reaction with reverse transcription (RT-PCR) from total larval RNA. BLAST searches and tree building suggested that *EtZic* is most closely related to *CsZic2* (ref. 21), *AmphiZic* (ref. 22), and the vertebrate *Zic2* (refs 18–20) genes. *In situ* hybridization with a digoxigenin-labelled probe was carried out as described previously³⁰.

Received 23 July; accepted 25 August 2004; doi:10.1038/nature02975.

- Hall, B. K. *The Neural Crest in Development and Evolution* (Springer, New York, 1999).
- Le Douarin, N. & Kalchauer, C. *The Neural Crest* 2nd edn (Cambridge Univ., Cambridge, 1999).
- Baker, C. V. H. & Bronner-Fraser, M. The origins of the neural crest. Part II: an evolutionary perspective. *Mech. Dev.* **69**, 13–29 (1997).
- Gans, C. & Northcutt, R. G. Neural crest and the origin of vertebrates: a new head. *Science* **220**, 268–274 (1983).
- Shimeld, S. M. & Holland, P. W. H. Vertebrate innovations. *Proc. Natl Acad. Sci. USA* **97**, 4449–4452 (2000).
- Wada, H. Origin and evolution of the neural crest: A hypothetical reconstruction of its evolutionary history. *Dev. Growth Diff.* **43**, 509–520 (2001).
- Stone, J. R. & Hall, B. K. Latent homologues of the neural crest as an evolutionary novelty. *Evol. Dev.* **6**, 123–129 (2004).
- Corbo, J. C., Erives, A., DiGregorio, A. & Levine, M. Dorsal/ventral patterning of the vertebrate neural tube is conserved in a protochordate. *Development* **124**, 2335–2344 (1997).
- Langeland, J. A., Tomsa, J. M., Jackman, W. R. & Kimmel, C. B. An amphioxus snail gene: Expression in paraxial mesoderm and neural plate suggests a conserved role in patterning the chordate embryo. *Dev. Genes Evol.* **208**, 569–577 (1998).
- Holland, L. Z. & Holland, N. D. Evolution of the neural crest and placodes: amphioxus as a model system. *J. Anat.* **199**, 85–98 (2001).
- Jeffery, W. R. & Swalla, B. J. Evolution of alternate modes of development in ascidians. *BioEssays* **14**, 219–226 (1992).
- Berrill, N. J. Studies in tunicate development. III. Differential retardation and acceleration. *Phil. Trans. R. Soc. Lond. B* **225**, 255–326 (1935).
- Lyerla, T. A., Lyerla, J. H. & Fisher, M. Pigmentation in the orange tunicate *Ecteinascidia turbinata*. *Biol. Bull.* **149**, 178–185 (1975).
- Thiery, J.-P., Duband, J. L. & Dolouvie, A. Pathways and mechanisms of avian trunk neural crest cell migration and localization. *Dev. Biol.* **93**, 324–343 (1982).
- Jungalwala, F. B., Chou, D. K. H., Suzuki, Y. & Maxwell, G. D. Temporal expression of HNK-1 reactive sulfolipid in cultured quail trunk neural crest cells—comparison with other developmentally regulated glycolipids. *J. Neurochem.* **58**, 1045–1051 (1992).
- Akitaya, T. & Bronner-Fraser, M. Expression of cell adhesion molecules during initiation and cessation of neural crest cell migration. *Dev. Dyn.* **194**, 12–20 (1992).
- Hirata, M., Ito, K. & Tsunek, K. Migration and colonization patterns of HNK-1 immunoreactive neural crest cells in lamprey and swordfish embryos. *Zool. Sci.* **14**, 305–312 (1997).
- Nakata, K., Nagai, T., Aruga, J. & Mikoshiba, K. *Xenopus* Zic family and its role in neural and neural crest development. *Mech. Dev.* **75**, 43–51 (1998).
- Brewster, R., Lee, J. & Ruiz-Altaba, A. Gli/Zic factors pattern the neural plate by determining domains of cell differentiation. *Nature* **398**, 579–583 (1998).
- Elms, P., Siggers, P., Napper, D., Greenfield, A. & Arkell, R. Zic2 is required for neural crest formation and hindbrain patterning during mouse development. *Dev. Biol.* **264**, 391–406 (2003).
- Atou, Y. *et al.* macho-1 related genes in *Ciona* embryos. *Dev. Genes Evol.* **212**, 87–92 (2002).
- Gostling, N. J. & Shimeld, S. M. Protochordate Zic genes define primitive compartments and highlight molecular changes underlying neural crest evolution. *Evol. Dev.* **5**, 136–144 (2003).
- Katz, M. J. Comparative anatomy of the tunicate tadpole *Ciona intestinalis*. *Biol. Bull.* **164**, 1–27 (1983).
- Wada, H., Saiga, H., Satoh, N. & Holland, P. W. H. Tripartite organization of the ancestral chordate brain and the antiquity of the placodes: Insights from ascidian Pax-2/5/8, Hox, and Otx genes. *Development* **125**, 1113–1122 (1998).
- Erickson, C. A. From the crest to the periphery: Control of pigment cell migration and lineage segregation. *Pigment. Cell Res.* **6**, 336–347 (1993).
- Burighel, P., Lane, N. J., Zaniolo, G. & Manni, L. The neurogenic role of the neural gland in the development of the ascidian *Botryllus schlosseri* (Tunicata, Ascidacea). *J. Comp. Neurol.* **394**, 230–241 (1998).
- Manni, L., Lane, N. J., Sorrentino, M., Zaniolo, G. & Burighel, P. Mechanisms of neurogenesis during the embryonic development of a tunicate. *J. Comp. Neurol.* **412**, 527–541 (1999).
- Manni, L., Lane, N. J., Burighel, P. & Zaniolo, G. Are neural crest and placodes exclusive to the vertebrates? *Evol. Dev.* **3**, 297–298 (2001).
- Thorndyke, M. C. & Probert, L. Calcitonin-like cells in the pharynx of the ascidian *Styela clava*. *Cell Tissue Res.* **203**, 301–309 (1979).
- Olsen, C. L. & Jeffery, W. R. A *forkhead* gene related to *HNF-3 β* is required for gastrulation and axis formation in the ascidian embryo. *Development* **124**, 3609–3619 (1997).

Acknowledgements This work was supported by grants to W.R.J. from the National Science Foundation, the National Institutes of Health, and the Bermuda Biological Station. We thank D. Martasian, A. Parkhurst, and L. Reed for technical assistance.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.R.J. (jeffery@umd.edu). The EtZic DNA sequence has been deposited in GenBank under accession number AY176563.

Role for a cortical input to hippocampal area CA1 in the consolidation of a long-term memory

Miguel Remondes & Erin M. Schuman

Caltech/HHMI, Division of Biology, 114-96, Pasadena, California 91125, USA

A dialogue between the hippocampus and the neocortex is thought to underlie the formation, consolidation and retrieval of episodic memories^{1–4}, although the nature of this cortico-hippocampal communication is poorly understood. Using selective electrolytic lesions in rats, here we examined the role of the direct entorhinal projection (temporoammonic, TA) to the hippocampal area CA1 in short-term (24 hours) and long-term (four weeks) spatial memory in the Morris water maze. When short-term memory was examined, both sham- and TA-lesioned animals showed a significant preference for the target quadrant. When re-tested four weeks later, sham-lesioned animals exhibited long-term memory; in contrast, the TA-lesioned animals no longer showed target quadrant preference. Many long-lasting memories require a process called consolidation, which involves the exchange of information between the cortex and hippocampus^{3,5,6}. The disruption of long-term memory by the TA lesion could reflect a requirement for TA input during either the acquisition or consolidation of long-term memory. To distinguish between these possibilities, we trained animals, verified their spatial memory 24 hours later, and then subjected trained animals to TA lesions. TA-lesioned animals still exhibited a deficit in long-term memory, indicating a disruption of consolidation. Animals in which the TA lesion was delayed by three weeks, however, showed a significant preference for the target quadrant, indicating that the memory had already been adequately consolidated at the time of the delayed lesion. These results indicate that, after learning, ongoing cortical input conveyed by the TA path is required to consolidate long-term spatial memory.

To assess the role of the TA pathway in the acquisition and retention of spatial memory in the rat, we identified stereotaxic coordinates to target the TA axons for electrolytic ablation. We used histology, electrophysiology, and retrograde tracing techniques to verify the extent and specificity of the TA lesions. We included in our behavioural analysis data from animals with relatively restricted lesions in the stratum lacunosum moleculare, the region where TA axons terminate in area CA1 (Fig. 1a,b and Supplementary Figs 1 and 2); we excluded animals in which the lesion extended significantly into the perforant path input to the dentate gyrus. We performed electrophysiological recordings on a subset of slices, and assessed synaptic transmission in the perforant path–dentate gyrus, TA–CA1 and CA1–subiculum pathways with extracellular stimulation and recordings (Fig. 1c). Brain slices from sham-lesioned animals showed normal synaptic transmission at all three sets of synapses (Fig. 1c). Slices from TA-lesioned animals exhibited

normal synaptic transmission at the perforant path–dentate gyrus and CA1–subiculum synapses, but a complete loss of synaptic responses at the TA–CA1 synapse (Fig. 1c). The TA axons originate in entorhinal cortex (EC) layer III and the trisynaptic pathway originates in EC layer II. Thus, using a unique set of coordinates, we injected a fluorescence-labelled retrograde tracer (cholera toxin subunit B) targeting both the dentate gyrus and CA1 and examined the distribution of fluorescence in the EC layers (Fig. 1d). Sham-lesioned animals exhibited significant fluorescence labelling in both layers II and III, whereas TA-lesioned animals exhibited significant labelling in layer II only (Fig. 1d). Taken together, these data indicate that we achieved a restricted ablation of TA input to CA1 without significant damage to the parallel trisynaptic circuit.

We tested the ability of the TA-lesioned rats to acquire, retain, consolidate and retrieve spatial information in the Morris water maze¹, comparing the performance of sham-, TA- and hippocampal-lesioned animals (Fig. 2a). Animals that received lesions did not differ significantly from sham-lesioned animals in their ability to navigate towards a visible platform (Supplementary Fig. 3). As shown by others, the complete ablation of the hippocampus prior to training resulted in significant learning impairments¹, as indicated by longer latencies (Fig. 2b) and reduced target quadrant preference during probe trials (Supplementary Fig. 3b). In contrast, animals

that received lesions restricted to the TA path showed immediate and short-term (24-h) spatial learning that was indistinguishable from sham-lesioned controls (Fig. 2c). Taken together with the anatomical data (Fig. 1), these data suggest that the hippocampal output is not substantially affected in the TA-lesioned animals. This result indicates that the TA input to the hippocampus is not required for rats to learn the platform location or for short-term memory.

Memories can be transformed from labile (short-term) to a relatively stable (long-term) form by a process referred to as consolidation. Some theories of long-term memory posit that following initial learning, a prolonged interaction between hippocampal and cortical circuits is required to consolidate the memory at the 'systems' level^{3,5,6}. We reasoned that the TA input might be important for these post-training cortico-hippocampal interactions. We first tested whether control animals exhibited long-term memory, assessed four weeks after training, for the target quadrant. In the absence of any additional (for example, beyond 24 h) exposure to the pool, sham-lesioned animals exhibited significant preference for the target quadrant four weeks after training (Fig. 2d), indicating that our training protocol is sufficient to

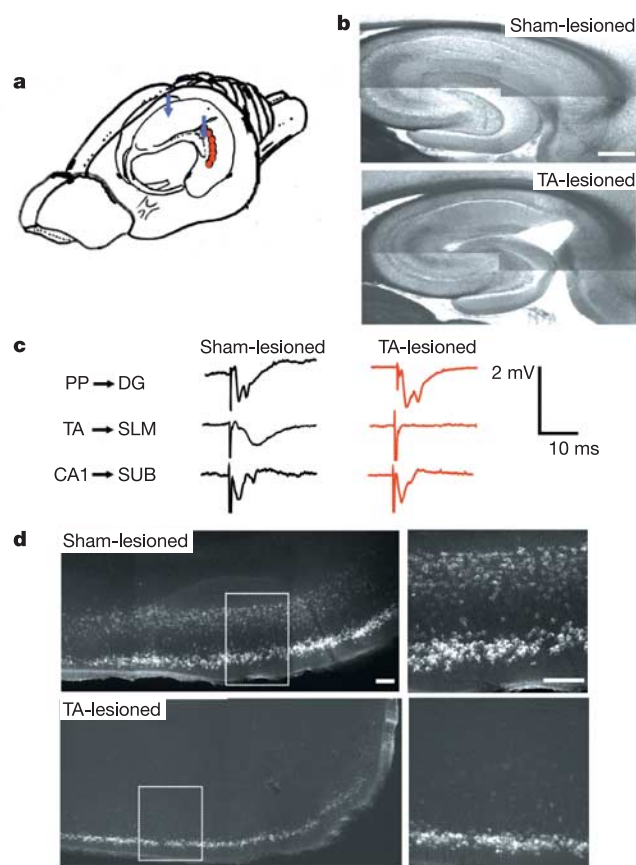


Figure 1 Specificity of TA lesions. **a**, Rat brain hippocampal formation with electrolytic lesion sites^{15,17,18} shown in red. **b**, Representative sections from sham- and TA-lesioned animals. Scale bar, 250 μ m. See also Supplementary Figs 1 and 2. **c**, Representative field potential recordings from hippocampal slices prepared from sham- or TA-lesioned animals. PP, perforant path; DG, dentate gyrus; SLM, stratum lacunosum moleculare; SUB, subiculum. **d**, Representative images of entorhinal cells labelled by the fluorescent retrograde tracer (injection positions shown by blue arrows in **a**). In sham-lesioned animals the tracer labelled cells in layers II and III. In TA-lesioned animals the labelling was largely confined to layer II. Scale bar, 100 μ m.

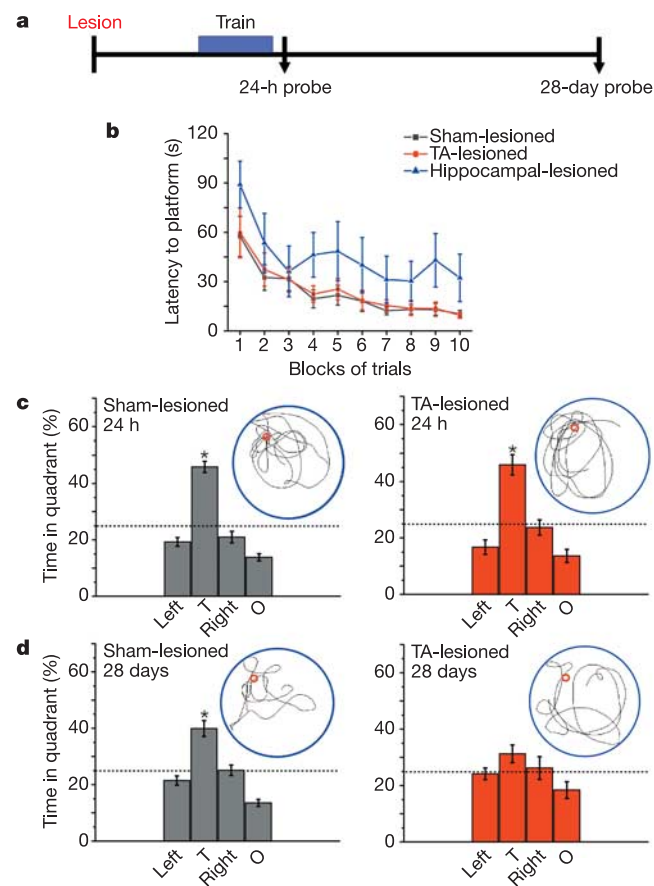


Figure 2 TA-lesioned animals exhibit deficits in long-term, but not short-term spatial memory. **a**, Experiment time-line. **b**, Mean escape latencies for sham-lesioned ($n = 21$), hippocampal-lesioned ($n = 5$) and TA-lesioned groups ($n = 10$). **c**, Mean time that each group spent in indicated quadrants during a short-term memory probe trial (24 h after training ended). Quadrants are left, T (target), right and O (opposite to target). Both groups showed a significant preference ($P \leq 0.05$) for the target quadrant. Error bars depict the standard error of the mean, s.e.m. Insets are a representative swim path from a single animal during a probe trial. **d**, Mean time that each group spent in each quadrant during a long-term memory probe trial (28 days after training ended). Sham-lesioned animals (15/21) still showed a significant preference for the target ($P \leq 0.05$) whereas, on average, TA-lesioned animals (4/10) no longer showed significant memory for the target.

produce both short-term (24 h) and long-term (four weeks) memory. Surprisingly, the TA-lesioned animals that exhibited significant target preference when tested after 24 h lost their memory for the target quadrant when tested four weeks after training (Fig. 2d). Taken together, these data indicate that TA lesions selectively affect long-term (~1 month) memory, without affecting short-term (~24 h) memory. These results suggest that the TA input to area CA1 is required for the establishment of a long-term spatial memory.

The disruption of long-term memory by the TA lesion could reflect a requirement for TA input to acquire long-term memory, or, alternatively, could reflect a requirement for TA-conveyed cortical input to consolidate a long-term spatial memory. To distinguish between these two possibilities, we lesioned the TA pathway after training and testing animals for short-term (24 h) memory (Fig. 3a). Before the administration of lesions, all animals learned the platform location at the same rate (Fig. 3b). In addition, both the sham- and future TA-lesioned animals showed significant preference for the target quadrant when examined 24 h after the end of training (Fig. 3c). Trained animals then received either a sham or a TA lesion; four weeks later, their preference for target quadrant was

re-assessed. Sham-lesioned animals still exhibited a significant preference for the target quadrant (Fig. 3d). TA-lesioned animals, however, no longer showed memory for the target quadrant (Fig. 3d). Because the lesion was administered after successful learning and short-term memory, the impaired performance of the lesioned animals indicates an ongoing requirement for TA input for the consolidation and/or retrieval of a long-term spatial memory, rather than its acquisition.

If the TA-conveyed cortical activity is required for consolidation of memory then a finite window of vulnerability to perturbation is predicted. To address this, we conducted an additional set of experiments in which the TA lesion was executed three weeks after learning; sham- and TA-lesioned animals were then given a final probe trial at 30 days (Fig. 4a). As before, prior to the administration of lesions, all animals learned the platform location at the same rate (Fig. 4b). In addition, both the sham- and future TA-lesioned animals showed a significant preference for the target quadrant when examined 24 h after the end of training (Supplementary Fig. 3c). Three weeks later, another probe trial was conducted and a randomly selected subset of the animals that exhibited significant target quadrant preference (Fig. 4c) received TA lesions. Nine days later (on day 30), a final probe trial revealed that both the sham- and the TA-lesioned animals exhibited a

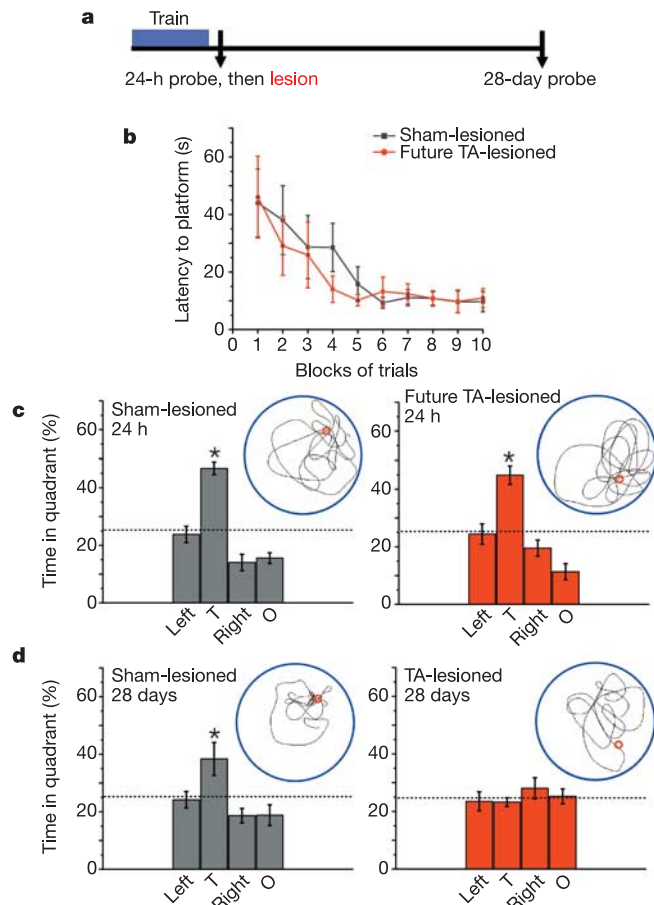


Figure 3 Disruption of the TA path 24 h after learning resulted in deficits in long-term spatial memory. **a**, Experiment time-line. **b**, Mean escape latencies for the sham-lesioned ($n = 8$) and future TA-lesioned ($n = 8$) groups, which showed similar learning rates. **c**, Mean time that each group spent in indicated quadrant during a short-term memory probe trial. Both groups showed a significant preference ($P \leq 0.05$) for the target quadrant. **d**, Mean time that each group spent in each quadrant during a long-term memory probe trial. Sham-lesioned animals (7/8) still showed a significant preference for the target quadrant ($P \leq 0.05$), whereas the TA-lesioned animals (0/8) no longer exhibited memory for the trained location. The groups differed significantly in their target preference ($P \leq 0.05$).

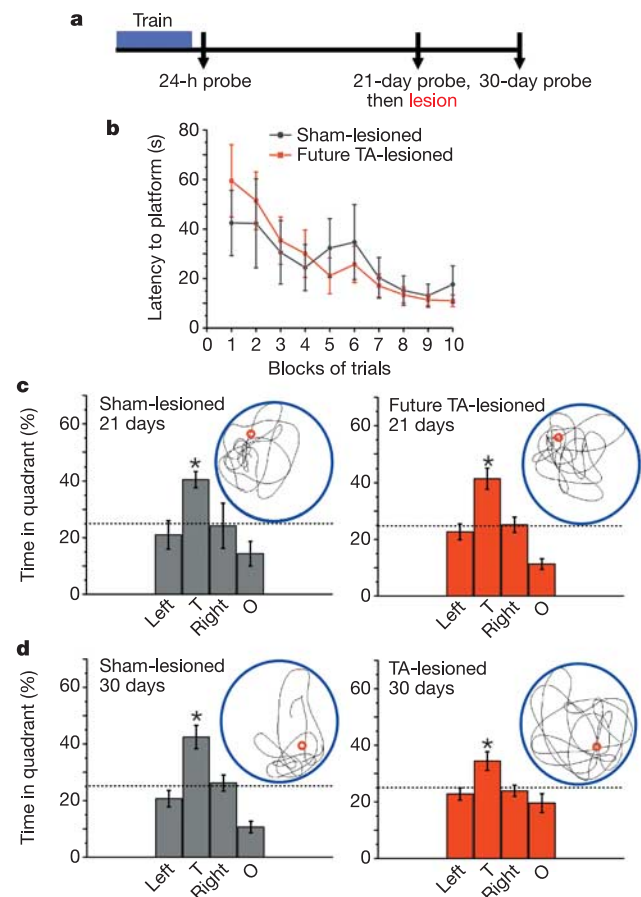


Figure 4 Disruption of the TA path three weeks after learning does not abolish long-term spatial memory. **a**, Experiment time-line. **b**, Mean escape latencies for the sham-lesioned ($n = 6$) and future TA-lesioned ($n = 9$) groups, which showed similar learning rates. **c**, Mean time that each group spent in indicated quadrant during a probe trial conducted 21 days after training. Both groups showed a significant preference ($P \leq 0.05$) for the target. **d**, Mean time that sham- or TA-lesioned animals spent in each quadrant during a long-term memory probe trial. Both sham-lesioned (6/6) and TA-lesioned animals (7/9) showed significant preference for the target quadrant ($P \leq 0.05$) and did not differ significantly from one another.

significant preference for the target quadrant relative to the other quadrants (Fig. 4d). Thus, a disruption of the TA input three weeks after learning failed to disrupt the long-term memory for the learned quadrant, whereas a TA disruption within 24 h (Fig. 3) abolished the memory. These results indicate that the temporal window for long-term memory consolidation lasts, on average, less than three weeks following training.

Previous studies have shown that deafferentation^{7,8} or ablation^{1,9} of the hippocampus disrupts the rate of learning and short-term memory for the platform location in the Morris water maze. Restricted lesions, however, that target individual pathways in the trisynaptic circuit have often resulted in little memory impairment. For example, lesions restricted to the dentate gyrus¹⁰ or area CA3^{11,12} do not affect the acquisition of spatial memory tasks^{11,12} or the development and stability of place cells in area CA1^{10,12}. In addition, mice possessing a conditional deletion of the NMDA receptor in area CA3, which abolishes commissural–CA3 long-term potentiation, show normal acquisition of spatial memory in the standard Morris water maze¹³. Our results suggest that the normal spatial learning exhibited by animals with targeted disruptions of the trisynaptic circuit might be due to information conveyed by the TA input to area CA1.

Recent studies have shown that TA–CA1 synapses exhibit synaptic plasticity, including LTP^{14–16} and LTD¹⁷. Appropriately timed TA bursting can also modulate Schaffer collateral-elicited spiking activity¹⁸ and plasticity¹⁵ in the hippocampal CA1 field. Transmission at the TA–CA1 can also be regulated by neuromodulators¹⁹ that are known to influence spatial memory^{20,21}. These observations provide a potential synaptic framework for understanding the regulation of memory consolidation by the TA input.

Thus our data show that animals that received selective lesions of the TA input to the hippocampus exhibited normal learning and short-term spatial memory, but failed to maintain the memory for an extended period of time. In a related experiment, pharmacologic inactivation of the hippocampus either 1–7 or 7–12 days after water maze training disrupted the memory for the target quadrant preference²². In addition, mice heterozygous for an α -CAMKII mutation show short-term memory for the target quadrant that decays over a 10–17-day retention period²³. In our experiments, a long-term memory deficit was evident when the TA input was interrupted before or 24 h after the learning experience, but not when disrupted three weeks after training, suggesting that the consolidation period had ended by this time. The observation that TA-lesioned animals were able to retrieve information about the correct target 24 h after learning, or when the TA-lesion was delayed to 21 days after training, argues against a simple deficit in retrieval. These results indicate that, following learning, ongoing cortical input conveyed by the TA path is required to consolidate long-term spatial memory. Coordinated activity between hippocampal and cortical circuits has been observed during sleep episodes^{24,25} often associated with memory consolidation^{26,27}. Further research should aim to establish the precise nature and function of the TA-conveyed cortical information during memory consolidation. □

Methods

Standard techniques used for surgical procedures, lesion execution and assessment and electrophysiology are described in the Supplementary Methods.

Animals

All animals used were male adult Long Evans rats (290–340 g; Charles River or Simonsen). Use of animals was performed according to the guidelines of the Caltech Institutional Animal Care and Use Committee.

Morris water maze

A pool 2.0 m in diameter filled with water made opaque with white non-toxic ink (Jazz Gloss Tempra white, Van Aken International) maintained at $25.0 \pm 2.0^\circ\text{C}$ was used. The

maze was located in a laboratory room, with distal cues at several locations. After a recovery period of at least ten days after surgery, animals were brought to the behaviour room (where they were housed for the duration of the training), handled for 1–2 days, and trained.

The full protocol consisted of seven days in which the first trial was always a probe trial, that is, there was no platform in the pool, to assess any spatial bias before the initiation of training and to observe the acquisition of quadrant preference during the previous trials. In probe trials, the animal was released from the centre of the pool, with its head facing a different room wall every day; animals were allowed to swim in the pool for 60 s. The quadrant preference was assessed as the percentage of total time rats spent in each of the four quadrants. The first training day consisted of a probe trial followed by a 'visible platform' trial, in which the platform was indicated by a red/orange flag. After this, rats were given their first 'hidden platform' learning trial, during which they were allowed to rest on the platform for 30 s before being released from one of the pool's starting points (north, south, east or west). Animals were allowed 120 s to find the platform and allowed to stay there for 30 s; if animals did not find the platform in 120 s they were picked up from the water and put on the platform for 30 s.

A block of trials consisted of four trials corresponding to four different randomized release points. The subsequent four days consisted of one probe trial followed by two training blocks (a total of eight trials) with at least a two-hour interval between blocks. To compensate for possible extinction resulting from the daily probe trial, the first training block (after the probe trial), and this one only, began with the placement of the rat on the platform for 30 s. On the last training day, a single training block followed the usual probe trial. On the following day animals were given a single probe trial, 24 h after the last learning trial—this is referred to in the text as the 24-h probe. Therefore, the total training took six days followed by the last 24-h probe on the seventh day. The rats were then returned to their home cages and were tested later for long-term memory 28–30 days after the 24-h probe trial. In two different sets of experiments, animals were lesioned either immediately after the 24-h probe trial, or after a 21-day-delay probe trial.

Data were collected by an HVS (UK) system—a charge-coupled device (CCD) camera connected to a PC card inserted in an IBM-compatible laptop computer. Software tracked the successive positions of the rat in the pool at a rate of 10 Hz for the duration of each trial. Data were analysed with the Water 2002 HVS (UK) software, and the summary data were transferred to Microcal Origin for further analysis and graphic display. The data for acquisition and retention were analysed by comparing the quadrant preferences and escape latencies averaged across animals of the same surgical group, using analysis of variance (ANOVA). The water maze probe data were analysed within group and experiment to assess quadrant preference, and later analysed between groups (ANOVA and unpaired *t*-test) to assess differences between quadrant preferences across different surgical and retention time groups. Differences were considered statistically significant at $P < 0.05$.

Received 8 June; accepted 16 August 2004; doi:10.1038/nature02965.

- Morris, R. G. M., Garrud, P., Rawlins, J. N. P. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
- Eichenbaum, H. A cortical-hippocampal system for declarative memory. *Nature Rev. Neurosci.* **1**, 41–50 (2001).
- Scoville, W. B. & Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neuropsychiatry Clin. Neurosci.* **12**, 103–113 (1957).
- Squire, L. R. & Zola, S. M. Amnesia, memory and brain systems. *Phil. Trans. R. Soc. Lond. B* **352**, 1663–1673 (1997).
- McGaugh, J. L. Time-dependent processes in memory storage. *Science* **153**, 1351–1358 (1966).
- Squire, L. R. & Alvarez, P. Retrograde amnesia and memory consolidation: a neurobiological perspective. *Curr. Opin. Neurobiol.* **5**, 169–177 (1995).
- Skeltton, R. W. & McNamara, R. K. Bilateral knife cuts to the perforant path disrupt spatial learning in the Morris water maze. *Hippocampus* **2**, 73–80 (1992).
- Eichenbaum, H., Stewart, C. & Morris, R. G. M. Hippocampal representation in place learning. *J. Neurosci.* **10**, 3531–3542 (1990).
- Sutherland, R. J., Whishaw, I. Q. & Kolb, B. A behavioural analysis of spatial localization following electrolytic, kainate- or colchicine-induced damage to the hippocampal formation in the rat. *Behav. Brain Res.* **7**, 133–153 (1983).
- McNaughton, B. L., Barnes, C. A., Meltzer, J. & Sutherland, R. J. Hippocampal granule cells are necessary for normal spatial learning but not for spatially-selective pyramidal cell discharge. *Exp. Brain Res.* **76**, 485–496 (1989).
- Jarrard, L. E. Selective hippocampal lesions and behaviour: effects of kainic acid lesions on performance of place and cue tasks. *Behav. Neurosci.* **97**, 873–889 (1983).
- Brun, V. H. et al. Place cells and place recognition maintained by the direct entorhinal-hippocampal circuitry. *Science* **296**, 2243–2246 (2002).
- Nakazawa, K. et al. Requirement for hippocampal CA3 NMDA receptors in associative memory recall. *Science* **297**, 211–218 (2002).
- Colbert, C. M. & Levy, W. B. Long-term potentiation of perforant path synapses in hippocampal CA1 in vitro. *Brain Res.* **606**, 87–91 (1993).
- Remondes, M. & Schuman, E. M. Direct cortical input modulates plasticity and spiking in CA1 pyramidal neurons. *Nature* **416**, 736–740 (2002).
- Remondes, M. & Schuman, E. M. Properties of early- and late-phase LTP at temporomnemonic-CA1 synapses. *Learning Memory* **10**, 247–252 (2003).
- Dvorak-Carbone, H. & Schuman, E. M. Patterned activity in stratum lacunosum moleculare inhibits CA1 pyramidal neuron firing. *J. Neurophys.* **82**, 3213–3222 (1999).
- Dvorak-Carbone, H. & Schuman, E. M. Long-term depression of temporomnemonic-CA1 hippocampal synaptic transmission. *J. Neurophys.* **81**, 1036–1044 (1999).
- Otmakhova, N. A. & Lisman, J. E. Dopamine selectively inhibits the direct cortical pathway to the CA1 hippocampal region. *J. Neurosci.* **19**, 1437–1445 (1999).
- El-Ghundi, M. et al. Spatial learning deficit in dopamine D1 receptor knockout mice. *Eur. J. Pharmacol.* **383**, 95–106 (1999).

21. Li, S., Cullen, W. K., Anwyl, R. & Rowan, M. J. Dopamine-dependent facilitation of LTP induction in hippocampal CA1 by exposure to spatial novelty. *Nature Neurosci.* **6**, 526–531 (2003).
22. Riedel, G. *et al.* Reversible neural inactivation reveals hippocampal participation in several memory processes. *Nature Neurosci.* **2**, 898–905 (1999).
23. Frankland, P. W., O'Brien, C. O., Ohno, M., Kirkwood, A. & Silva, A. J. Alpha-CAMKII-dependent plasticity in the cortex is required for permanent memory. *Nature* **411**, 309–313 (2001).
24. Siapas, A. G. & Wilson, M. A. Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron* **21**, 1123–1128 (1998).
25. Sirota, A., Csicsvari, J., Bhul, D. & Buzsáki, G. Communication between neocortex and hippocampus during sleep in rodents. *Proc. Natl Acad. Sci. USA* **100**, 2065–2069 (2003).
26. Wilson, M. A. & McNaughton, B. L. Reactivation of hippocampal ensemble memories during sleep. *Science* **265**, 676–679 (1994).
27. Lee, A. K. & Wilson, M. A. Memory of sequential experience in the hippocampus during slow wave sleep. *Neuron* **36**, 1183–1194 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Siapas, G. Laurent, M. Sutton and other members of the Schuman laboratory for discussions. This work was supported by the Fundacao para a Ciencia e Tecnologia (FCT)—Portugal and the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.M.S. (schumane@its.caltech.edu).

Enhanced virulence of influenza A viruses with the haemagglutinin of the 1918 pandemic virus

Darwyn Kobasa^{1*}, Ayato Takada^{2,3}, Kyoko Shinya^{2,4}, Masato Hatta¹, Peter Halfmann¹, Steven Theriault⁴, Hiroshi Suzuki⁵, Hidekazu Nishimura⁶, Keiko Mitamura^{7*}, Norio Sugaya^{7*}, Taichi Usui⁸, Takeomi Murata⁸, Yasuko Maeda², Shinji Watanabe¹, M. Suresh¹, Takashi Suzuki^{3,9}, Yasuo Suzuki^{3,9}, Heinz Feldmann⁴ & Yoshihiro Kawaoka^{1,2,3}

¹Department of Pathobiological Sciences, University of Wisconsin, Madison, Wisconsin 53706, USA

²Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan

³CREST, Japan Science and Technology Agency, Saitama, 332-0012, Japan

⁴Special Pathogens Program, National Microbiology Laboratory, Health Canada and Department of Medical Microbiology, University of Manitoba, Winnipeg, Manitoba R3E 3R2, Canada

⁵Department of Public Health, School of Medicine, Niigata University, Niigata, 951-8510, Japan

⁶Viral Research Center Clinical Research Division, Sendai National Hospital, Sendai, 983-8520, Japan

⁷Department of Pediatrics, Nippon Kogan Hospital, Kawasaki, 210-0852, Japan

⁸Department of Applied Biological Chemistry, Shizuoka University, Shizuoka 422-8529, Japan

⁹Department of Biochemistry, University of Shizuoka, School of Pharmaceutical Sciences and COE Program in the 21st Century, Yada, Shizuoka-shi 422-8526, Japan

* Present addresses: National Microbiology Laboratory, Canadian Science Centre for Human and Animal Health, Winnipeg, Manitoba R3E 3R2, Canada (D.K.); Institute for Animal Experimentation, Tohoku University Graduate School of Medicine, Sendai 980-8575, Japan (K.S.); Department of Pediatrics, Kawasaki Municipal Hospital, Kawasaki 210-0013, Japan (K.M.); Department of Pediatrics, Keiyu Hospital, Yokohama 220-0012, Japan (N.S.)

The 'Spanish' influenza pandemic of 1918–19 was the most devastating outbreak of infectious disease in recorded history. At least 20 million people¹ died from their illness, which was characterized by an unusually severe and rapid clinical course. The complete sequencing of several genes of the 1918 influenza virus has made it possible to study the functions of the proteins

encoded by these genes in viruses generated by reverse genetics, a technique that permits the generation of infectious viruses entirely from cloned complementary DNA. Thus, to identify properties of the 1918 pandemic influenza A strain that might be related to its extraordinary virulence, viruses were produced containing the viral haemagglutinin² (HA) and neuraminidase³ (NA) genes of the 1918 strain. The HA of this strain supports the pathogenicity of a mouse-adapted virus in this animal^{4,5}. Here we demonstrate that the HA of the 1918 virus confers enhanced pathogenicity in mice to recent human viruses that are otherwise non-pathogenic in this host. Moreover, these highly virulent recombinant viruses expressing the 1918 viral HA could infect the entire lung and induce high levels of macrophage-derived chemokines and cytokines, which resulted in infiltration of inflammatory cells and severe haemorrhage, hallmarks of the illness produced during the original pandemic⁶.

The sequence of the HA receptor/fusion gene of the 1918 influenza A virus (H^{SP}) was determined in 1999 (ref. 2), followed shortly by the NA gene (N^{SP})³, which codes for the viral sialidase. These molecules, which represent the major viral surface glycoproteins of influenza virus, are important targets of the host immune response, and critical changes in their structure can expand the viral host range and enhance the virulence of infection^{7,8}. However, neither protein of the 1918 virus contains sequence motifs known to be associated with high virulence^{2,3}. Attempts to account for the unusually high virulence of the virus have focused on the living conditions of soldiers, a highly susceptible group, and of civilians at the end of the First World War⁹, as well as on potentially unique properties of the virus itself^{10,11}. The combination of H^{SP} and N^{SP} in an A/WSN/33 (WSN) (H1N1) genetic background was shown to be pathogenic in mice without prior adaptation of the proteins to growth in these hosts^{4,5}, usually a prerequisite for producing a lethal infection in mice with a human virus. However, WSN itself is both adapted to and highly pathogenic in mice, so it was not possible to assess the possible contribution that H^{SP} and N^{SP} had made to the pathogenicity of the recombinant virus. Another candidate, the non-structural gene of the 1918 virus, had an attenuating effect when tested in a recombinant virus bearing the genetic background of A/Puerto Rico/8/34 (H1N1), also a highly pathogenic virus in mice¹⁰. Viruses generated with the matrix, nucleoprotein and non-structural protein genes of WSN replaced by the respective genes from the 1918 virus, together with H^{SP} and N^{SP}, were also pathogenic in mice^{4,5} but not significantly more than the recombinant possessing only the H^{SP} and N^{SP} genes. Thus, the molecular basis for the unprecedented virulence of the 1918 pandemic virus remains uncertain, opening the way for studies that directly test the contribution of H^{SP} and N^{SP} to pathogenicity.

We therefore used reverse genetics¹² to produce a panel of recombinant viruses containing H^{SP} together with N^{SP} or the NA gene of A/Kawasaki/173/2001 (H1N1) (K173) in the background of the remaining negative-sense RNA gene segments derived from three influenza viruses: WSN, K173 and A/Memphis/8/88 (H3N2) (M88) (Table I). All recombinant viruses formed plaques on Madin–Darby canine kidney (MDCK) cell monolayers in the presence of trypsin and replicated efficiently to a high titre (data not shown).

To assess the contribution of H^{SP} and N^{SP} to pathogenicity *in vivo*, mice were intranasally inoculated with the recombinant and parental viruses. Each virus possessing H^{SP} as well as the WSN virus replicated to a high titre in the lungs by day 3 post-infection (p.i.; Table I) and was pathogenic, resulting in serious morbidity and eventually death. Notably, the K173 and M88 parental human viruses did not cause discernible morbidity in mice and replicated to only low titres in the lungs, whereas the K173/H^{SP} and M88/H^{SP} viruses, in which the original HA proteins were replaced with H^{SP}, caused lethal infection in mice. The N^{SP} protein did not contribute to viral pathogenicity, as viruses containing H^{SP} and N^{SP} were not

more pathogenic than those with the K173 NA. By contrast, the HA and NA proteins of WSN did not significantly enhance the virulence of either M88 or K173, as indicated by the lack of illness in mice inoculated with recombinant M88/HN^{WSN} and limited illness associated with K173/HN^{WSN} infection. The ability of WSN/H^{SP}N^{SP} (H^{SP} and N^{SP} in the WSN background) and WSN to replicate in organs outside the respiratory tract was also assessed, with no evidence of replication detected in the heart, spleen, liver, kidneys or brain. Intracerebral inoculation of mice with 6×10^5 plaque-forming units (PFU) of WSN, WSN/H^{SP}N^{SP} or K173/H^{SP}N^{SP} (H^{SP} and N^{SP} in the K173 background) led to replication of only WSN in the brain (data not shown), an outcome consistent with the capacity of plasminogen binding by the NA to activate the HA and increase the neurotropism of WSN¹³, properties not shared by the NA protein of either the K173 or the 1918 virus. The relative pathogenicities of the viruses bearing H^{SP} depended on the type of virus providing the other gene segments. For WSN/H^{SP}N^{SP} the dose required to kill 50% of infected mice (MLD₅₀) was similar to that of WSN (Table 1). MLD₅₀ values for the other viruses were higher than for WSN/H^{SP}N^{SP}, a reflection of the contribution of genes from the mouse-adapted WSN compared to the non-adapted human viruses. Thus, it appears that H^{SP} possesses unknown properties, not found in the WSN HA, that determine its ability to enhance the virulence of influenza A viruses in mice.

In a further effort to understand the increased virulence of the 1918 virus, we examined the lungs of mice intranasally infected with 10 MLD₅₀ (see Methods for details of MLD₅₀ calculation) of WSN, WSN/H^{SP}N^{SP}, M88/H^{SP}N^{SP} (H^{SP} and N^{SP} in the M88 background) or M88/H^{SP}, or a dose of M88 that was equivalent (in PFU) to 10 MLD₅₀ of M88/H^{SP}N^{SP} (as M88 was not lethal). By day 6 p.i. the viral antigen distribution for WSN/H^{SP}N^{SP}, M88/H^{SP}N^{SP} and M88/H^{SP} was considerably wider—encompassing the bronchus, bronchioli and almost the entire alveolar epithelium—than that seen for M88 or even WSN (Fig. 1a–c; see also Supplementary Figs 1 and 2; data for M88/H^{SP}N^{SP} not shown). Infection with viruses possessing H^{SP} was characterized by massive recruitment of polymorphonuclear cells (mainly neutrophils) accompanied by intra-alveolar haemorrhage (Fig. 1d; see also Supplementary Figs 1 and 2). M88, on the other hand, produced limited inflammatory foci that resulted from initial infection of the bronchioli with only minor extensions into the adjacent alveoli. Most of the lung had a similar appearance to normal, uninfected controls (Fig. 1e, f; see also Supplementary Fig. 1) and only limited viral antigen was detected (Fig. 1; see also Supplementary Fig. 1). The presence of H^{SP} in M88/H^{SP} permitted greater viral replication over a longer period compared with M88, as shown by detection of viral antigen on day 6 p.i. (Fig. 1a, b), but the pathogenicity of viruses containing H^{SP}

was not related simply to the extent of viral replication. High titres of virus were recovered from the lungs of mice infected with both WSN and WSN/H^{SP}N^{SP} on both days 3 and 6 p.i., but lesions produced by WSN/H^{SP}N^{SP} in lung tissue were clearly more severe (Supplementary Table 1 and Supplementary Fig. 2a, b). The lungs of mice infected with WSN had the characteristics of bronchopneumonia, but inflammatory reactions and viral antigens were confined mainly to the bronchus, bronchioli and peribronchiolar alveoli. As the M88/H^{SP} recombinant, which possesses only the HA from the 1918 virus, caused severe lung infection similar to that seen with WSN/H^{SP}N^{SP} or M88/H^{SP}N^{SP} infection, these results indicate that H^{SP} was the primary determinant of enhanced pathogenicity in the recombinant viruses. Severe lung infection was a hallmark of the illness produced by the original pandemic virus in humans⁶, suggesting a possible, although not conclusive, association between the pandemic virus HA and its pathogenicity in humans, and indicating the need to examine the contribution of the HA to viral pathogenicity in other animal models.

The severity of influenza-virus-induced disease in humans correlates with the ability of virulent viral strains to induce pro-inflammatory cytokines in macrophages^{14,15}. Additionally, there is a strong association between cytokine levels and severity of clinical symptoms in people infected with influenza^{16–18}. Therefore it is possible that pathogenicity of H^{SP}-expressing viruses might be causally related to their enhanced ability to induce pro-inflamma-

Table 1 Properties of recombinant viruses

Virus	Origin of genes*			MLD ₅₀ †	Lung titre‡
	HA	NA	Others		
WSN/H ^{SP} N ^{SP}	1918	1918	WSN	3.0	5.0 ± 0.1
WSN	WSN	WSN	WSN	3.3	5.0 ± 0.3
M88/H ^{SP} N ^{SP}	1918	1918	M88	5.2	4.7 ± 0.2
M88/H ^{SP}	1918	K173	M88	4.4	5.1 ± 0.1
M88	M88	M88	M88	>6.2	2.9 ± 0.2
M88/HN ^{WSN}	WSN	WSN	M88	>6.7	ND
K173/H ^{SP} N ^{SP}	1918	1918	K173	6.9	ND
K173/H ^{SP}	1918	K173	K173	5.2	4.8 ± 0.2
K173	K173	K173	K173	>7.4	3.5 ± 0.3
K173/HN ^{WSN}	WSN	WSN	K173	>6.9	ND

ND, no data.

* Recombinant viruses were generated with gene segments from the 1918 pandemic virus, A/WSN/33 (WSN), A/Memphis/8/88 (M88) and A/Kawasaki/173/2001 (K173).

† The virus dose required to kill 50% of mice (MLD₅₀) was determined by intranasally inoculating mice with the test virus, as described in Methods.

‡ Average (±s.d.) lung titres in three mice studied on day 3 p.i. expressed as log₁₀ PFU per g lung tissue.

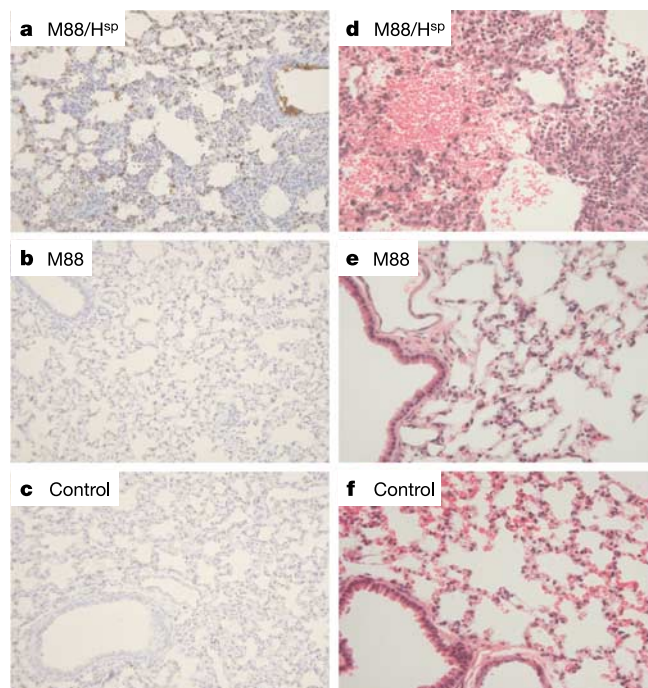


Figure 1 Pathological examination of lungs infected with the M88/H^{SP} or M88 viruses. **a**, Viral antigens were distributed throughout the entire lung of mice infected with M88/H^{SP}, including the bronchiole and alveoli of all lobes. **b**, By contrast, in lung tissue of mice infected with only the background virus M88, viral antigens were not detected on day 6 p.i. **d**, Global severe inflammatory reactions extending from the bronchus to the bronchiole to the alveoli were prominent in the lungs of mice infected with M88/H^{SP}. Numerous neutrophilic infiltrations were prominent in the peribronchial area and alveolar wall. Haemorrhage and oedematous lesions were common. These observations contrast with the lungs of mice infected with M88, which produced only small, irregularly distributed inflammatory lesions that were self-limited. Indeed, the vast majority of the lung was not affected (**e**), similar to normal lung (**f**). **a, d**, M88/H^{SP}, lung, 6 days p.i.; **b, e**, M88, lung, 6 days p.i.; **c, f**, control, lung, 6 days p.i. **a–c**, Immunostaining with a rabbit polyclonal antibody to WSN; **d–f**, haematoxylin and eosin staining.

tory cytokines and inflict immune damage to tissues in infected mice. To probe this possibility, we compared levels of cytokines/chemokines in lungs between viral isolates that differed in their virulence. Notably, the levels of monocyte chemotactic protein 1 (MCP-1), macrophage inflammatory protein (MIP)-1 β , MIP-2, MIP-3 α , interleukin (IL)-1 β , IL-6, IL-12 (p40), IL-18 and granulocyte colony stimulating factor (G-CSF) in the lungs of M88/H^{SP}-infected mice were substantially higher than in M88-infected mice (Fig. 3). Because activated macrophages are the major

source of IL-1 β , IL-6, IL-12, IL-18 and G-CSF during inflammation, the presence of high levels of these cytokines is strongly indicative of macrophage activation (Supplementary Fig. 3). Whereas the chemokines MCP-1, MIP-1 β and MIP-3 α are capable of inducing mixed leukocyte recruitment to the site of inflammation, the strong induction of MIP-2 (a strong chemoattractant for neutrophils) is highly consistent with the infiltration of granulocytes in the lungs of M88/H^{SP}-infected mice. Taken together, these data suggested that H^{SP} is a critical determinant of macrophage activation, especially early on in infection (see day 1 data in Fig. 3), and of production of chemoattractant for neutrophils, which subsequently leads to the trafficking of neutrophils and acute lung injury^{19,20}.

The 1918 virus is thought to have originated from an avian source². Hence, the relative ability of the HA to recognize the preferred receptor species for human viruses—sialic acid bound to a neighbouring galactose by an α 2,6 linkage (NeuAc α -2,6Gal) in contrast to the preferred avian virus receptor, NeuAc α -2,3Gal²¹—is thought to affect its ability to infect humans and cause disease. Using a competitive binding assay, we showed that the H^{SP} preferentially recognizes NeuAc α -2,6Gal (Fig. 2). Although humans possess both types of receptor species in respiratory epithelia and are susceptible to infection with viruses bearing HA proteins of either receptor specificity²², all of the human viruses studied thus far preferentially recognize NeuAc α -2,6Gal over NeuAc α -2,3Gal²¹. Thus, our data suggest that the virus that eventually became the 1918 pandemic strain must have circulated in humans, or possibly another adaptive host (such as pigs) with human-type receptors, for a sufficient time to develop a preference for the human sialic acid receptor. This model gains support from the recently solved crystal structure of H^{SP} (refs 23, 24), which predicts that the protein would preferentially bind to the human receptor²³.

An additional factor influencing the severity of an influenza pandemic or epidemic is the level of pre-existing immunity to the

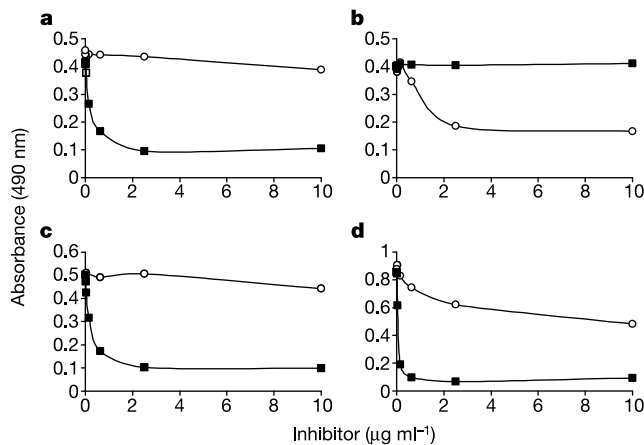


Figure 2 Competitive binding assay for haemagglutinin receptor specificity. **a–d**, The relative binding affinities of HA in viruses WSN/H^{SP}N^{SP} (**a**), A/dk/HK/836/80 (**b**), A/Kawasaki/173/01 (**c**) and A/New Caledonia/20/99 (**d**) were determined by competitive inhibition of HRP-conjugated fetuin binding by NeuAc α -2,3Gal (open circles) or NeuAc α -2,6Gal (filled squares) sialylglycopolymers.

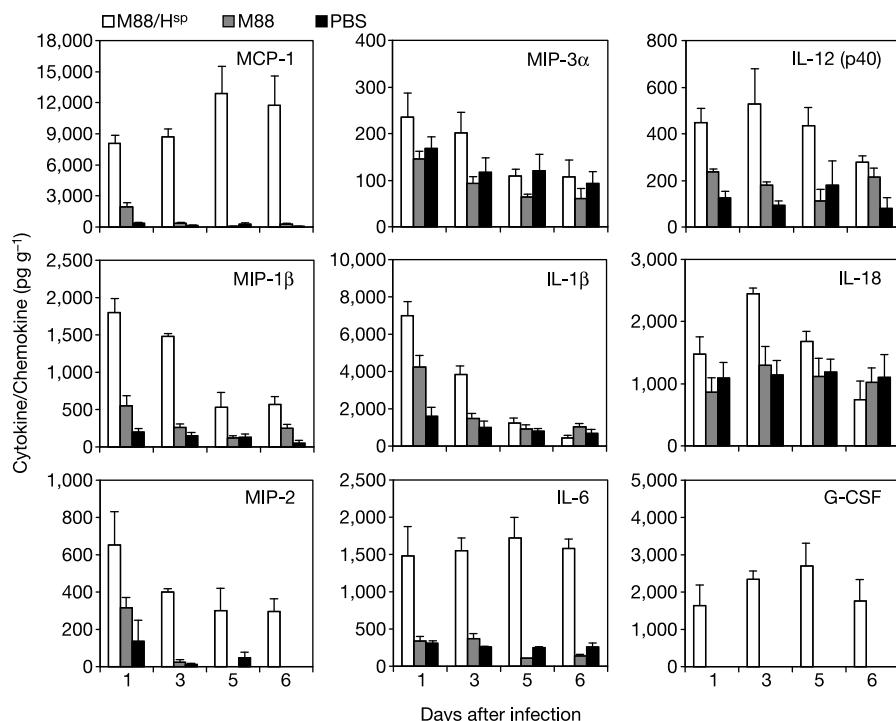


Figure 3 Cytokine expression in lungs of mice infected with M88/H^{SP} and M88 or mock infected (PBS). The expression of cytokines and chemokines in lung homogenates was determined by ELISA on days 1, 3, 5 and 6 p.i. Data are presented as the mass

of cytokine (pg) per g of lung tissue ($\pm$ standard deviation). (See also Supplementary Fig. 3.)

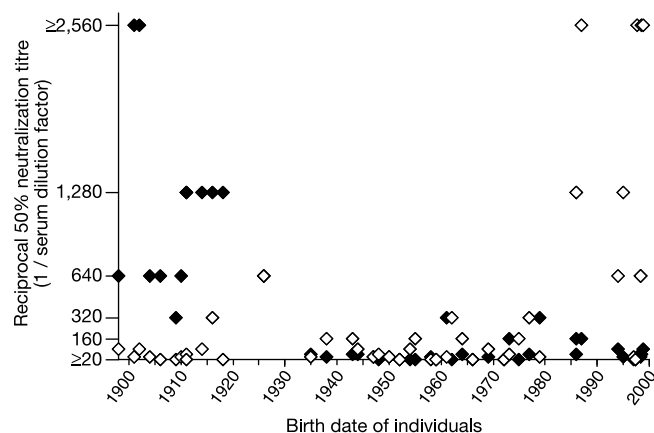


Figure 4 Neutralization activities in human sera against a virus possessing H^{sp} and N^{sp}. Neutralization titres of sera from a broad cross-section of donors were measured against WSN/H^{sp}N^{sp} (filled diamonds) and a recent human H1N1 virus (open diamonds) in a plaque reduction assay. Donors younger than 14 yr were recently infected with an H1N1 virus, with the exception of two cases of H3N2 infection. The history of influenza infection in other donors was unknown. The reciprocal titre of serum required to neutralize at least 50% of each virus (50% neutralizing titre) is plotted against the date of birth of the patients.

virus. To assess the extent to which the human population would be protected from the emergence of a 1918-like virus, we compared the neutralizing activities in a panel of human sera obtained in 1999 from patients of widely differing ages. The majority of the serum donors younger than 15 yr of age, most of whom had recently been infected with an H1N1 virus, showed strong neutralizing activity against a recent H1N1 test virus, but only limited activity against the WSN/H^{sp}N^{sp} virus (Fig. 4; see also Supplementary Table 2). By contrast, serum samples from donors who were alive in 1918 showed markedly high activities against the WSN/H^{sp}N^{sp} virus compared with those against recent H1N1 and H3N2 isolates. Donors born between 1934 and 1978 had only limited neutralizing activity against the WSN/H^{sp}N^{sp} virus. Thus, a large segment of the population would be susceptible to an outbreak of a 1918-like influenza virus. The only group with significant natural protection would be survivors of the 1918 pandemic, who still express high levels of antibodies against an antigen to which they were exposed to over 80 yr ago, a phenomenon referred to as original antigenic sin^{25,26}.

Analysis of sequence information derived directly from viral genes in the preserved tissues of people infected during the 1918 pandemic offers unparalleled opportunity to assess the contributions of the HA, NA and other viral proteins to the extreme virulence of this unique pathogen. Using a reverse genetics technique to generate a panel of reassortants containing different combinations of the H^{sp} and N^{sp} genes in varied genetic backgrounds, we show that the HA protein (but not the NA protein) from the 1918 virus is essential for highly virulent infection in an animal model. However, the virulence of influenza virus is probably a polygenic trait, in that ultimately we expect other gene products, such as the NS1 protein²⁷, to be implicated in the phenotype of the 1918 virus. Sequence analysis of the H^{sp} gene by others suggests that it originated in birds² and therefore is likely to be contained in viruses still circulating in the avian pool, where viral proteins remain relatively unchanged. Once the properties of the H^{sp} gene that gave rise to its lethal infectivity are better understood, it should be possible to devise effective control measures and to improve global surveillance networks for influenza viruses that pose the greatest threat to humans as well as other animal species.

Methods

Viruses

Full-length HA (GenBank accession number AF117241) and NA (GenBank AF250356) genes from the 1918 influenza A virus were constructed by using successive polymerase chain reaction amplifications with 70-mer oligonucleotides and cloned into a plasmid vector (pPoll) containing the RNA polymerase I promoter and terminator to generate pPoll-1918HA and pPoll-1918NA. Recombinant viruses WSN/H^{sp}N^{sp}, K173/H^{sp}N^{sp}, K173H^{sp}, M88/H^{sp}N^{sp}, M88/H^{sp}, K173/HN^{WSN}, M88/HN^{WSN} as well as parental viruses K173, M88 and WSN were generated by reverse genetics as previously described¹². Viruses were passaged twice on MDCK cells, in the presence of 1.0 µg ml⁻¹ TPCK-treated trypsin in minimal essential medium (MEM) supplemented with 0.3% bovine serum albumin and antibiotics. Influenza viruses A/New Caledonia/20/99 (H1N1) (NewCal), A/duck/Hong Kong/836/80 (H3N1) (dk/HK/80) and another recent clinical isolate, A/Kawasaki/233/2001 (H1N1) (K233), were similarly propagated in MDCK cells. All procedures related to viruses containing genes from the 1918 virus were performed in a Biosafety level 4 facility at the National Microbiology Laboratory, Health Canada, in Winnipeg, or in enhanced BSL3 at the University of Wisconsin-Madison.

In vitro analysis of viral replication

Semi-confluent MDCK cells were infected at a multiplicity of infection of 0.001 PFU per cell with each virus maintained under the same growth conditions used for the propagation of viruses. The supernatant was sampled at times 0, 15, 24, 39 and 48 h p.i., and titres were determined by standard plaque assay on confluent MDCK cells in the presence of 1.0 µg ml⁻¹ TPCK-trypsin. The results of duplicate assays are reported.

HA receptor analysis by solid-phase binding assay

Viruses were allowed to attach overnight at 4 °C in wells of 96-well plates as previously described²⁸. Polymer-bound conjugates containing the sialic acid *N*-acetylneuraminic acid linked to galactose through either an α-2,3 or α-2,6 bond (Neu5Acα2,3LacNAc-pAP and Neu5Acα2,6LacNAc-pAP) were prepared as previously described²⁹. Serial dilutions of each sialylglycoconjugate were prepared in phosphate-buffered saline (PBS) containing 0.01% Tween-20 (TPBS), and 25 µl was added to the wells containing the immobilized viruses together with 25 µl of 47 nM horseradish peroxidase (HRP)-conjugated bovine fetuin in TPBS. After incubation at 4 °C for 2 h, the wells were extensively washed with PBS, and 100 µl per well of *o*-phenylenediamine (0.4 mg ml⁻¹) in 0.02% H₂O₂, 50 mM citrate-phosphate buffer, pH 5.5 was added. After incubation at 25 °C for 15 min, the absorbance at 490 nm was read.

Viral pathogenicity and lethal dose determination

Isoflurane-anesthetized 6-week-old female BALB/C mice were intranasally inoculated with tenfold serial dilutions (three mice per dilution) of viruses in 100 µl of PBS. The mice were monitored daily for 21 days for disease symptoms and survival. The dose required to kill 50% of mice (MLD₅₀) was calculated by the method of ref. 30.

To determine the extent of viral replication in mice, we intranasally inoculated animals with 10 MLD₅₀ of M88/H^{sp}N^{sp}, M88/H^{sp}, K173/H^{sp}N^{sp}, K173/H^{sp}, WSN or WSN/H^{sp}N^{sp}, or with a dose of non-pathogenic virus, M88 or K173, equivalent in PFU to 10 MLD₅₀ of M88/H^{sp}N^{sp} and K173/H^{sp}N^{sp}, respectively. On days 3 and 6 p.i., the lungs, heart, spleen, kidneys, liver and brain were collected from mice infected with WSN and WSN/H^{sp}N^{sp} (*n* = 5). On day 3 p.i., the lungs of mice infected with the remaining viruses were collected. The organs were homogenized, and virus growth was determined in each tissue by plaque assay on MDCK cells.

To assess the ability of viruses possessing the 1918 virus HA and NA genes to replicate in the central nervous system, we intracranially inoculated 4-week-old BALB/C mice with 6 × 10⁵ PFU of WSN, WSN/H^{sp}N^{sp} or K173/H^{sp}N^{sp}. Brains were collected on day 4 p.i. and homogenized, and virus replication was determined by plaque assay on MDCK cells.

For pathological analysis mice were infected with 10 MLD₅₀ of the M88, M88/H^{sp}N^{sp}, M88/H^{sp}, WSN and WSN/H^{sp}N^{sp} viruses. After mice were killed on days 3 and 6 p.i., the lungs were removed and fixed in 10% phosphate-buffered formalin. The tissues were then dehydrated, embedded in paraffin, and cut into 5-µm-thick sections that were stained with standard haematoxylin and eosin. For viral antigen detection, sections were processed for immunostaining by the two-step dextran polymer method (DAKO), with a rabbit polyclonal antibody to WSN used as the primary antibody.

All animal experiments were performed under an approved animal use document and according to the guidelines of the Canadian Council on Animal Care.

Cytokine/chemokine analysis

Isoflurane-anesthetized 6-week-old BALB/C mice were intranasally inoculated with 10 MLD₅₀ of M88/H^{sp} or 1.5 × 10⁶ PFU M88. On days 1, 3, 5 and 6 p.i., lungs were collected from three mice infected with each virus and homogenized in PBS. MCP-1, MIP-1α, MIP-1β, MIP-2, MIP-3α, IL-1α, IL-1β, IL-6, IL-10, IL-12 (p40), IL-12 (p70), IL-18, G-CSF, granulocyte-macrophage (GM)-CSF, tumour-necrosis factor (TNF)-α, IP-10, RANTES, vascular endothelial growth factor (VEGF), interferon (IFN)-α, and IFN-γ were detected by enzyme-linked immunosorbent assay (ELISA) (R&D Systems).

Human serum neutralization assays

Serum samples were collected between 1998–2001 from donors whose ages ranged from 15 months to 102 yr. The sera were treated with receptor-destroying enzyme (RDE)

(Accurate Chemical and Scientific Corp.) to destroy inhibitors of influenza virus replication. After inactivation of the RDE by treatment at 56 °C for 1 h, WSN/H³N^{SP} and K233 were each incubated with twofold serial dilutions of each serum at 37 °C for 1 h. Remaining infectivity was determined by titration of the samples in a plaque assay on MDCK cells.

Received 28 July; accepted 10 August 2004; doi:10.1038/nature02951.

- Johnson, N. P. & Mueller, J. Updating the accounts: global mortality of the 1918–1920 “Spanish” influenza pandemic. *Bull. Hist. Med.* **76**, 105–115 (2002).
- Reid, A. H., Fanning, T. G., Hultin, J. V. & Taubenberger, J. K. Origin and evolution of the 1918 “Spanish” influenza virus hemagglutinin gene. *Proc. Natl Acad. Sci. USA* **96**, 1651–1656 (1999).
- Reid, A. H., Fanning, T. G., Janczewski, T. A. & Taubenberger, J. K. Characterization of the 1918 “Spanish” influenza virus neuraminidase gene. *Proc. Natl Acad. Sci. USA* **97**, 6785–6790 (2000).
- Tumpey, T. M. *et al.* Existing antivirals are effective against influenza viruses with genes from the 1918 pandemic virus. *Proc. Natl Acad. Sci. USA* **99**, 13849–13854 (2002).
- Tumpey, T. M. *et al.* Pathogenicity and immunogenicity of influenza viruses with genes from the 1918 pandemic virus. *Proc. Natl Acad. Sci. USA* **101**, 3166–3171 (2004).
- Oxford, J. S. Influenza A pandemics of the 20th century with special reference to 1918: virology, pathology and epidemiology. *Rev. Med. Virol.* **10**, 119–133 (2000).
- Lamb, R. A. & Krug, R. M. *Fields Virology* (eds Knipe, D. M. *et al.*) 1487–1531 (Lippincott-Raven, Philadelphia, Pennsylvania, 2001).
- Wright, P. F. & Webster, R. G. *Fields Virology* (eds Knipe, D. M. *et al.*) 1533–1579 (Lippincott-Raven, Philadelphia, Pennsylvania, 2001).
- Oxford, J. S. *et al.* World War I may have allowed the emergence of “Spanish” influenza. *Lancet Infect. Dis.* **2**, 111–114 (2002).
- Basler, C. F. *et al.* Sequence of the 1918 pandemic influenza virus nonstructural gene (NS) segment and characterization of recombinant viruses bearing the 1918 NS genes. *Proc. Natl Acad. Sci. USA* **98**, 2746–2751 (2001).
- Chen, W. *et al.* A novel influenza A virus mitochondrial protein that induces cell death. *Nature Med.* **7**, 1306–1312 (2001).
- Neumann, G. *et al.* Generation of influenza A viruses entirely from cloned cDNAs. *Proc. Natl Acad. Sci. USA* **96**, 9345–9350 (1999).
- Goto, H., Wells, K., Takada, A. & Kawaoka, Y. Plasminogen-binding activity of neuraminidase determines the pathogenicity of influenza A virus. *J. Virol.* **75**, 9297–9301 (2001).
- Guan, Y. *et al.* H5N1 influenza: a protean pandemic threat. *Proc. Natl Acad. Sci. USA* **101**, 8156–8161 (2004).
- Cheung, C. Y. *et al.* Induction of proinflammatory cytokines in human macrophages by influenza A (H5N1) viruses: a mechanism for the unusual severity of human disease? *Lancet* **360**, 1831–1837 (2002).
- Hayden, F. G. *et al.* Local and systemic cytokine responses during experimental human influenza A virus infection. Relation to symptom formation and host defense. *J. Clin. Invest.* **101**, 643–649 (1998).
- Fritz, R. S. *et al.* Nasal cytokine and chemokine responses in experimental influenza A virus infection: results of a placebo-controlled trial of intravenous zanamivir treatment. *J. Infect. Dis.* **180**, 586–593 (1999).
- Kaiser, L., Fritz, R. S., Straus, S. E., Gubareva, L. & Hayden, F. G. Symptom pathogenesis during acute influenza: interleukin-6 and other cytokine responses. *J. Med. Virol.* **64**, 262–268 (2001).
- Abraham, E. Neutrophils and acute lung injury. *Crit. Care Med.* **31** (4 suppl.), S195–S199 (2003).
- Lee, W. L. & Downey, G. P. Neutrophil activation and acute lung injury. *Curr. Opin. Crit. Care* **7**, 1–7 (2001).
- Connor, J. S., Kawaoka, Y., Webster, R. G. & Paulson, J. C. Receptor specificity in human, avian and equine H2 and H3 influenza virus isolates. *Virology* **15**, 17–23 (1994).
- Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A. & Klenk, H. D. Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc. Natl Acad. Sci. USA* **101**, 4620–4624 (2004).
- Gamblin, S. J. *et al.* The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303**, 1838–1842 (2004).
- Stevens, J. *et al.* Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* **303**, 1866–1870 (2004).
- Francis, T. Jr Influenza: The new acquaintance. *Ann. Intern. Med.* **39**, 203–221 (1953).
- Fazekas De St Groth, S. & Webster, R. G. Disquisitions on original antigenic sin. I. Evidence in man. *J. Exp. Med.* **124**, 331–345 (1966).
- Geiss, G. K. *et al.* Cellular transcriptional profiling in influenza A virus-infected lung epithelial cells: the role of the nonstructural NS1 protein in the evasion of the host innate defense and its potential contribution to pandemic influenza. *Proc. Natl Acad. Sci. USA* **99**, 10736–10741 (2002).
- Gambaryan, A. S. & Matrosovich, M. N. A solid-phase enzyme-linked assay for influenza virus receptor-binding activity. *J. Virol. Methods* **39**, 111–123 (1992).
- Totani, K. *et al.* Chemoenzymatic synthesis and application of glycopolymers containing multivalent sialyloligosaccharides with a poly(L-glutamic acid) backbone for inhibition of infection by influenza viruses. *Glycobiology* **13**, 315–326 (2003).
- Reed, L. J. & Muench, H. A simple method of estimating fifty per cent endpoints. *Am. J. Hyg.* **27**, 493–497 (1938).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Dick, A. Grolla, M. Garbutt and S. Jones for assistance with BL4 procedures, M. McGregor and K. Wells for technical assistance, J. Gilbert for editing the manuscript, and Y. Kawaoka for illustrations. This work was supported by a grant-in-aid from the Japanese Ministry of Education, Culture, Sports, Science and Technology, CREST (Japan Science and Technology Corporation), NIAID Public Health Service research grants and Health Canada.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.K. (kawaoka@svm.vetmed.wisc.edu).

Hedgehog signalling in prostate regeneration, neoplasia and metastasis

Sunil S. Karhadkar^{1,2}, G. Steven Bova^{2–4}, Nadia Abdallah², Surajit Dhara², Dale Gardner⁵, Anirban Maitra², John T. Isaacs^{3,4}, David M. Berman^{1–4} & Philip A. Beachy^{1,4}

¹Department of Molecular Biology and Genetics and the Howard Hughes Medical Institute, ²Departments of Pathology, ³Urology and ⁴Oncology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

⁵USDA ARS, Poisonous Plant Research Laboratory, Logan, Utah 84341, USA

Metastatic cancers adopt certain properties of normal cells in developing or regenerating organs, such as the ability to proliferate and alter tissue organization. We find here that activity of the Hedgehog (Hh) signalling pathway, which has essential roles in developmental patterning^{1–6}, is required for regeneration of prostate epithelium, and that continuous pathway activation transforms prostate progenitor cells and renders them tumorigenic. Elevated pathway activity furthermore distinguishes metastatic from localized prostate cancer, and pathway manipulation can modulate invasiveness and metastasis. Pathway activity is triggered in response to endogenous expression of Hh ligands, and is dependent upon the expression of Smoothened, an essential Hh response component^{1,2,7} that is not expressed in benign prostate epithelial cells. Monitoring and manipulating Hh pathway activity may thus offer significant improvements in diagnosis and treatment of prostate cancers with metastatic potential.

Hedgehog signalling influences development and homeostasis of many gut-derived organs, some of which give rise to Hh-dependent cancers. In the foregut, for example, ligand-dependent pathway activity is required for growth of a significant proportion of small-cell lung cancers and of carcinomas of the stomach, oesophagus, pancreas and biliary tract^{8–10}. We consider here the prostate, the site of origin for the second most lethal malignancy in men. The prostate derives from embryonic endoderm that is caudal to the intestinal portal, and which also gives rise to organs such as the colon. Although tumorigenesis in the colon invariably involves Wnt pathway activation, the prostate has recently been found to exhibit a developmental patterning role for Hh signalling^{4–6} similar to that in the lung³.

We examined expression of Hh pathway ligands and endogenous targets in human prostate cancer cell lines by measuring levels of messenger RNA that encode the pathway components *GLI* and *PATCHED* (*PTCH*). Both *GLI* and *PTCH* are transcriptional targets of pathway activation with opposite roles in pathway response; *GLI* serves as a positive transcriptional effector and *PTCH* functions to restrain pathway activity by suppressing the action of Smoothened

(SMO). This negative function of PTCH is blocked by the binding of Hh ligands, and the pathway can thus be activated through Hh-mediated or mutational inactivation of PTCH; either form of pathway activation requires SMO (reviewed in refs 1, 2, 7).

All four prostate cancer cell lines examined (PC3, DU145, CWR22RV1 and LnCAP) expressed transcripts encoding *Sonic* (*SHH*) and *Indian* (*IHH*) *hedgehog* ligands (Fig. 1a). Tumour cells also expressed *PTCH* and *GLI* transcripts, and pathway activity was confirmed by quantitative real-time analysis of polymerase chain reaction with reverse transcription (RT-PCR), which revealed *PTCH* mRNA levels to be elevated ~200–400-fold in cancer cells, relative to benign prostate epithelial cells (PrE cells; see below) (Fig. 1b). Introduction of a Hh-responsive *GLI*-luciferase reporter¹¹ also revealed high luciferase activity in tumour cells, which was further augmented by addition of exogenous Shh ligand (ShhNp) and was blocked by treatment with a ligand-neutralizing monoclonal antibody (5E1) (ref. 8) (Supplementary Fig. 1a). Activity was also fully suppressible by treatment with cyclopamine, which specifically inhibits the Hh pathway response by binding to and stabilizing the inactive conformation of SMO^{11,12}. As seen in 22RV1-GLI cells, cyclopamine blockade of SMO was bypassed by stable overexpression of *GLI*, demonstrating the specificity of the cyclopamine effect in the Hh pathway.

Cyclopamine treatment inhibited growth of PC3, DU145 and 22RV1 cells (Fig. 1c) and expression of c-Myc and cyclin D1 (Supplementary Fig. 1c, d), as compared with treatment with the structurally related but inactive compound, tomatidine^{8,13}. This anti-proliferative effect of cyclopamine again was bypassed by overexpression of *GLI*, but not by an altered inactive form, *GLI*^{zfd}

(ref. 14) (Fig. 1c), and growth inhibition was confirmed in PC3 cells by treatment with 5E1 antibody (Supplementary Fig. 1b).

We further tested the role of Hh pathway activity *in vivo* by treating established subcutaneous PC3 and 22RV1 xenograft tumours in athymic mice with daily subcutaneous injections of cyclopamine (10 or 50 mg kg⁻¹) or of vehicle alone. We observed suppression of tumour growth at 10 mg kg⁻¹ cyclopamine, and actual regression at 50 mg kg⁻¹. Animals treated at the intermediate dose of 10 mg kg⁻¹ were killed and a 90% reduction in staining for the proliferation antigen Ki67 was noted (Supplementary Fig. 1f). Animals receiving treatment at 50 mg kg⁻¹ showed complete regression of the tumours within 20–24 days (d) of treatment (Fig. 1d, e; Supplementary Fig. 1g). Cessation of treatment did not result in re-growth of tumours, even after observation periods of 72 d (PC3) and 148 d (22RV1). Cyclopamine-treated 22RV1-GLI tumours grew faster than vehicle-treated 22RV1 tumours (Fig. 1e), and this acceleration of tumour growth by *GLI* overexpression even with cyclopamine treatment confirms *in vivo* that the rate of tumour cell growth corresponds to the degree of Hh pathway activity (Fig. 1c, e; Supplementary Fig. 1a).

Complete and durable tumour regression suggests that cells capable of renewing the tumour, that is, of functioning as tumour stem cells^{2,15,16}, require Hh pathway activity for their maintenance. In this regard, it is notable that cyclopamine suppressed transcription of the gene encoding nestin, an intermediate filament protein whose expression has not been described previously in the prostate, but which marks progenitor cell populations elsewhere in the endoderm as well as in neural tissues and muscle^{17–19} (Supplementary Fig. 1e). In addition to nestin, the Polycomb group protein

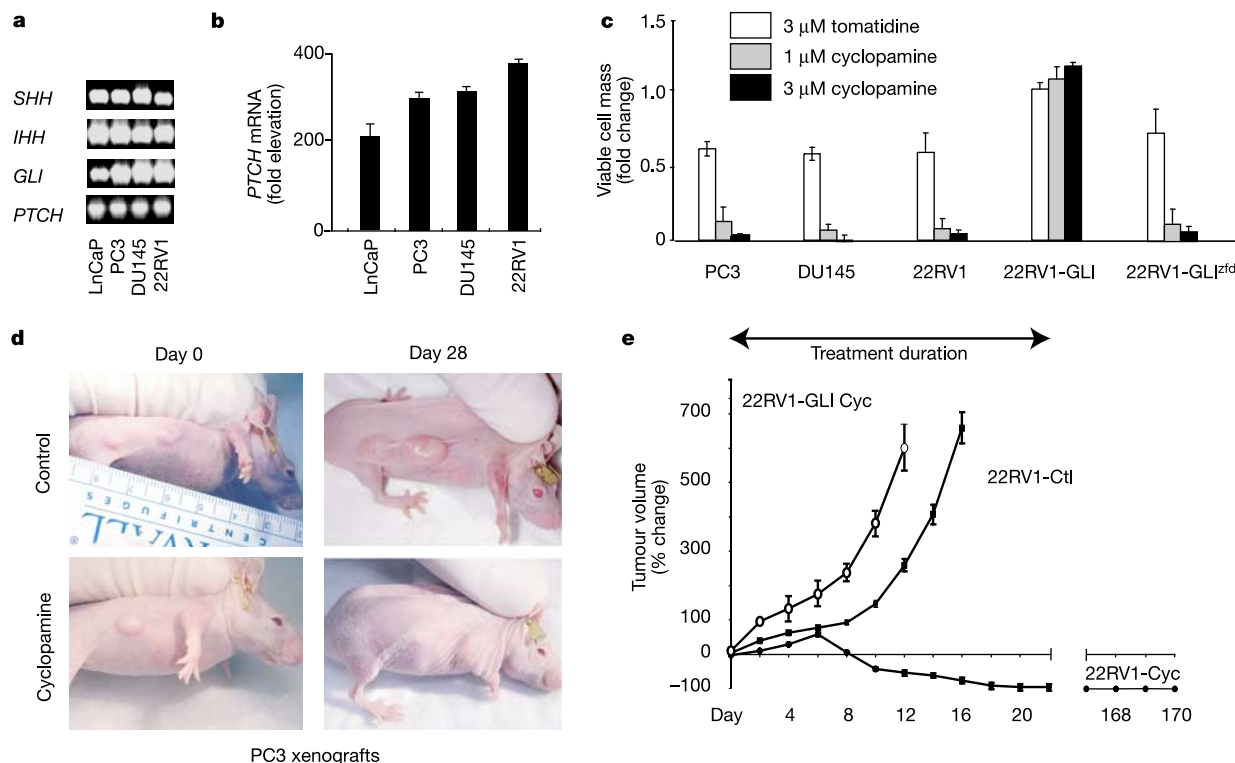


Figure 1 Hh pathway activity in growth of human prostate cancer cells. **a**, RT-PCR assay indicating expression of transcripts encoding IHH and SHH ligands and the Hh pathway targets *PTCH* and *GLI* in all prostate cancer cell lines examined. **b**, Quantitative real-time RT-PCR for *PTCH* mRNA relative to levels in benign prostate epithelial cells. **c**, Dose-dependent inhibition of growth in prostate cancer cells cultured with cyclopamine. 22RV1-GLI cells maintained Hh pathway activity and growth, whereas 22RV1-GLI^{zfd} cells remained susceptible. Results show the average of three

experiments. **d**, **e**, PC3 (**d**) and 22RV1 (**e**) prostate cancer xenograft tumours (median size 155 mm³) completely regressed after 22–28 d of cyclopamine treatment (50 mg kg⁻¹ d⁻¹) and did not recur during 58 (PC3) or 148 d (22RV1) of subsequent observation. Vehicle-treated tumours (22RV1-Ctl) and cyclopamine-treated 22RV1-GLI tumours grew rapidly. Error bars indicate standard error of the mean (s.e.m.)

Bmi-1, required for self-renewal of somatic stem cells in haematopoietic and neural lineages²⁰, is also expressed in a Hedgehog-pathway-dependent manner (data not shown).

To further investigate the role of Hh pathway activity in prostate progenitor cells, we examined epithelial regeneration in rodent ventral prostate using castration-induced androgen withdrawal to ablate prostate epithelium²¹. Seven days of androgen withdrawal markedly reduces epithelial content (by >90%) and is thought to leave a population greatly enriched in progenitor cells^{21,22}. Following a 10-day course of androgen supplementation (dihydrotestosterone; DHT, 50 mg kg⁻¹ d⁻¹), the ventral prostate regenerates to normal size (Fig. 2a) and histological appearance (that is, large glands lined with tall columnar epithelium; Fig. 2b). However, Hh pathway blockade by cyclopamine or Hh-neutralizing 5E1 antibody abolished prostate regeneration (Fig. 2a, b), yielding small glands lined with low cuboidal epithelium, similar in appearance to prostates in vehicle-treated castrates (Fig. 2b).

To address the potential relationship between tumour stem cells and epithelial progenitor cells we used a cultured cell system that

allows limited growth of cells that are capable of populating the epithelium of developing prostate glands (PrE cells) and that resemble prostate basal cells, a putative progenitor cell in prostate epithelium²³. We found that PrE cells stably transfected for expression of GLI (PrE-GLI cells), unlike mock transfected controls, showed unlimited growth *in vitro* (data not shown) and increased expression of mRNAs encoding c-Myc, cyclin D1 and nestin (Supplementary Fig. 2b). PrE-GLI cells furthermore formed aggressive subcutaneous tumours (Fig. 2c) and showed reactivity with antibodies against cytokeratins and Nkx3.1 (not shown), indicating a prostate epithelial phenotype. Hh pathway activity thus promotes growth of human prostate cancer cells and regeneration of prostate epithelium from endogenous progenitors, and can produce malignant transformation of progenitor-like primary cells when continuously induced.

Extension of our studies to lethal metastases removed at autopsy, as well as localized tumours and adjacent normal tissue from radical prostatectomies (Supplementary Tables 1 and 2), revealed that all samples of normal and localized or metastatic malignant prostate tissue expressed *SHH* and *IHH* (Fig. 3a), as did benign PrE cells. We found, however, that metastatic tumours ($n = 15$ samples from 12 patients) but not PrE cells nor benign prostate samples ($n = 12$ histologically verified normal tissue samples), expressed Hh pathway targets *PTCH* and *GLI* (Fig. 3a, b; Supplementary Fig. 3a, b). Only 3 of 12 samples from localized malignancies expressed *PTCH* and *GLI*, and quantitative real-time RT-PCR analysis (Fig. 3b) further revealed that *PTCH* mRNA levels in these three samples never exceeded one-tenth of that noted in the lowest-expressing metastatic tumours, indicating that Hh pathway activity is strongly correlated with prostate cancer metastasis.

Given marked differences in pathway activity, despite expression of Hh ligands in all benign and malignant prostate samples, we surveyed expression of positively acting components of Hh pathway response and found that SMO mRNA, although detectable in all metastatic cancer samples, was not detectable in non-responsive PrE cells, nor in normal prostate tissue (Fig. 3a; Supplementary Fig. 3a). We found that transfection for expression of SMO resulted in Hh reporter activity in PrE cells, which could be augmented by addition of exogenous ShhNp (Fig. 3c) and blocked by treatment with cyclopamine or 5E1 antibody, indicating that SMO expression confers the normal operation of a ligand-dependent Hh pathway response.

As human prostate cancer xenografts metastasize slowly and infrequently in mouse models, we explored the role of Hh pathway activity in metastasis using a series of rodent cell lines established from tumours with varying metastatic potential^{24,25}. Of six cell lines surveyed for Gli-luciferase reporter activity (Fig. 4a), the three characterized as highly metastatic (Mat-LyLu, AT3.1 and AT6.3) showed high levels of pathway activity relative to the three poorly metastatic lines (G, AT1 and AT2.1) (Fig. 4a; Supplementary Fig. 4a). By comparing a single cell line from the high (AT6.3) metastasis group with one from the low (AT2.1) metastasis group, we found that AT6.3 was particularly responsive to exogenous ShhNp ligand (Supplementary Fig. 4b), and that its growth was inhibited by cyclopamine and by 5E1 antibody (Supplementary Fig. 4c and data not shown). AT6.3 cells subcutaneously inoculated in nude mice extensively colonized visceral organs in the thoracic and abdominal cavities (Fig. 4b; Supplementary Fig. 4d), and mice bearing these tumours invariably died within a few weeks of inoculation (Fig. 4c). AT2.1 cells produced no mortality and no evidence of metastasis 30 d after subcutaneous inoculation (Fig. 4b, c), despite continued local growth.

We augmented Hh pathway activity by stable transfection of GLI in AT2.1 cells (AT2.1-GLI) and noted that all mice inoculated subcutaneously with these cells died within 16 d ($n = 6$), comparable to the 18-day maximal survival of mice ($n = 11$) inoculated with AT6.3 cells (Fig. 4c). AT2.1-GLI cells also produced widespread

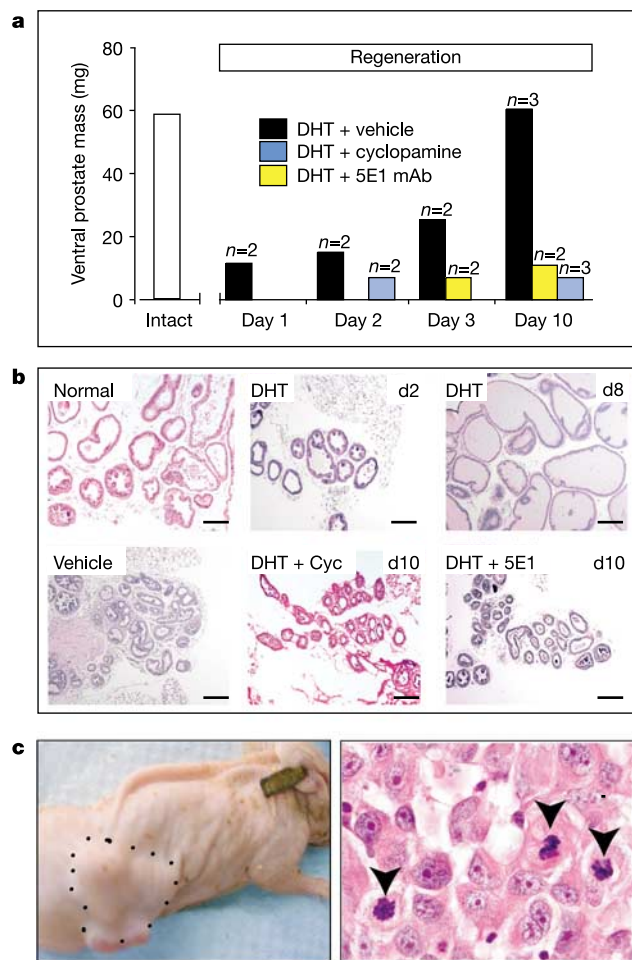


Figure 2 Hh pathway activity in prostate progenitors. **a**, Wet weights of ventral prostate glands decreased ~fivefold in male castrates and returned to normal within 10 d of androgen replacement with DHT (50 mg kg⁻¹ d⁻¹). *In vivo* Hh pathway blockade with cyclopamine or 5E1 mAb blocked prostate regeneration. **b**, Photomicrographs of control and regenerating ventral prostate glands. Glands from castrates treated with DHT and cyclopamine, or DHT and 5E1 mAb are smaller, similar to vehicle controls. Scale bar, 200 μm. **c**, Tumour formed by PrE-GLI cells in an athymic mouse (left, dashed lines). Photomicrograph (H&E stain, right) shows carcinoma cells with abnormal nuclei and abundant mitoses (arrowheads), characteristic of aggressive growth.

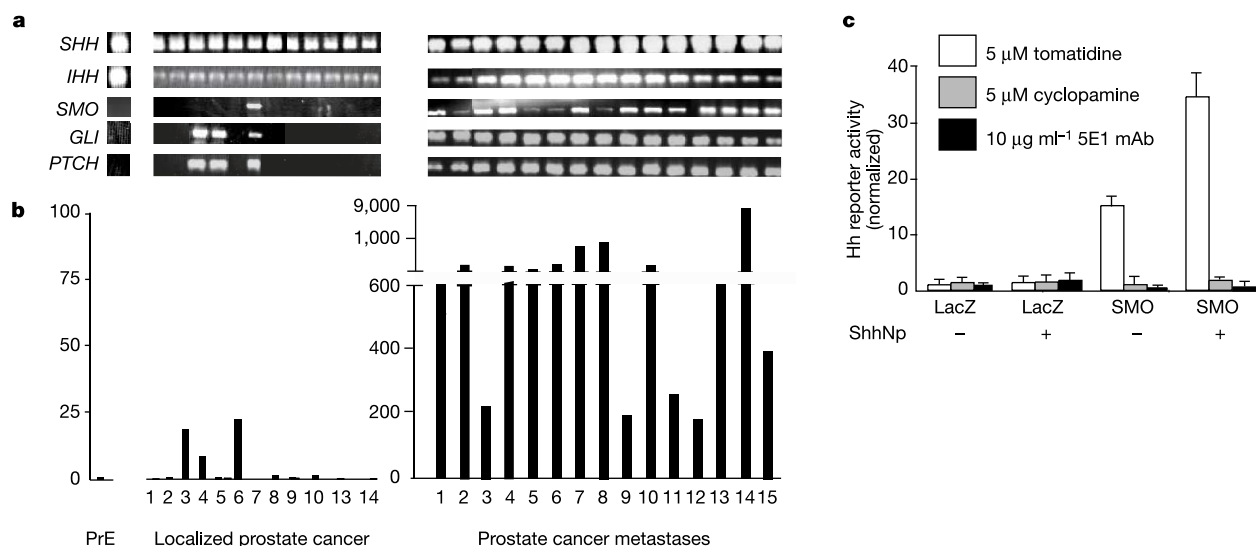


Figure 3 Hh pathway activation in metastatic prostate cancer and determination of Hh pathway responsiveness by SMO. **a**, RT-PCR assays indicating expression of IHH and SHH ligands in benign PrE cells (far left), and in localized ($n = 12$) and metastatic ($n = 15$) samples from 12 patients) prostate tumours. *PTCH* and *GLI* expression indicated Hh pathway activity in all metastases, but was undetectable in benign PrE cells or in benign prostate tissue (see Supplementary Fig. 3). Three of twelve localized tumours expressed detectable *PTCH* and *GLI*, and one of these also showed SMO expression. **b**, Quantitative

real-time RT-PCR shows *PTCH* levels in metastases $\geq$ tenfold higher than levels in localized tumours (note change of scale in y axis). **c**, Hh ligand responsiveness is conferred by SMO. Unlike control (LacZ) transfected PrE cells, cells transfected to express SMO exhibit endogenous Hh reporter activity that is augmented by exogenous ShhNp and abolished by cyclopamine or Hh neutralizing antibody (5E1 mAb). Error bars indicate s.e.m.

visceral metastases (Fig. 4b; data not shown), and activation of Hh pathway targets thus seems sufficient for conferral of a lethal metastatic phenotype. To determine whether the metastatic phenotype of AT6.3 cells could be reversed by Hh pathway blockade, we administered cyclopamine by intraperitoneal injection. This route of administration is less effective in providing continuous pathway inhibition (Supplementary Fig. 4e) and permitted growth of subcutaneous AT6.3 tumours, unlike subcutaneous administration (data not shown), but nevertheless improved survival to a median of 19 d at $10 \text{ mg kg}^{-1} \text{ d}^{-1}$, and inhibited metastasis and prevented death throughout a 50-day treatment period at $50 \text{ mg kg}^{-1} \text{ d}^{-1}$ (Fig. 4b, c; Supplementary Fig. 4d).

Although the primary AT6.3 subcutaneous tumours continued to grow under the 50 and $10 \text{ mg kg}^{-1} \text{ d}^{-1}$ intraperitoneal treatment regimens, the rate of growth was reduced from that of vehicle-treated tumours (Supplementary Table 3). In addition, conversion of AT2.1 to a metastatic phenotype by overexpression of GLI also increased growth rate, raising the possibility that growth rate may determine metastatic potential. We therefore measured the ability of cells to traverse a Matrigel-coated membrane with $8\text{-}\mu\text{m}$ pores and colonize the side of the membrane opposite that on which they are seeded (modified Boyden chamber assay), a correlate of metastatic potential *in vivo*^{26–28}. AT2.1-GLI and AT6.3 cells readily penetrated the matrix and populated the bottom surface of the membrane (the side opposite seeding), whereas AT6.3 cells treated with cyclopamine and AT2.1 cells rarely did so (Fig. 4d, e). The 22RV1 human cell line also exhibited cyclopamine-sensitive invasiveness, and cyclopamine effects in both human and rodent cells were bypassed by expression of Gli (Fig. 4e; Supplementary Fig. 4f). Growth rate of cells was not a significant factor in these assays owing to the short timescale (20 h) and to the normalization of invading cells to total viable cell mass.

Metastasis-associated invasiveness of epithelial tumours is thought to involve a transition from epithelial to mesenchymal character (EMT), associated with expression of the transcription factor Snail and consequent reduction in levels of proteins such as

E-cadherin that maintain epithelial organization²⁸. We found that Snail mRNA expression, strongly stimulated by GLI expression in AT2.1 cells, is constitutive in AT6.3 cells and can be suppressed by cyclopamine (Fig. 4f). The levels of E-cadherin mRNA are low in AT2.1-GLI and AT6.3 cells, and are highest in AT2.1 cells and cyclopamine-treated AT6.3 cells (Supplementary Fig. 4g), in correlation with low metastatic potential and reciprocal to the expression levels of Snail mRNA. The *Ndr1* gene, specifically associated with suppression of the metastatic phenotype without appreciable effects on proliferation in prostate and colon cancer^{27,29}, is also expressed in benign and non-metastatic tumour cells, but not in metastatic tumour cells unless Hh pathway blockade is imposed with cyclopamine (Supplementary Fig. 4h).

Approximately one in six human prostate cancers manifests the ability to metastasize and cause death. The occurrence of Hh pathway activity in all metastases and the ability of pathway activation to promote cell invasiveness, EMT and a metastasis-associated programme of gene expression suggest that pathway activity may be a prerequisite for metastasis. This activity could arise as indolent tumours progress or could be determined at the outset of tumour initiation³⁰ as a result of distinct initiating lesions affecting particular prostate cell types. The pathway activity evident in a minority of apparently localized prostate tumours (see above) might then indicate unrealized metastatic potential, a possibility that could be tested in prospective studies assessing the likelihood of metastasis as a function of pathway activity in prostate tumours.

Our findings that Hh pathway activity is required for regeneration of prostate epithelium, propagation of prostate cancer in xenografts and expression of the stem cell renewal factors nestin and Bmi-1 (refs 17–20) in cancer cells, and the additional finding that forced Hh pathway activity can produce malignant transformation of primitive prostate epithelial progenitor cells, together suggest that prostate cancer may be initiated by trapping of a normal stem cell in a Hh-dependent state of continuous renewal. The limiting factor for pathway activation seems to be SMO, suggesting that

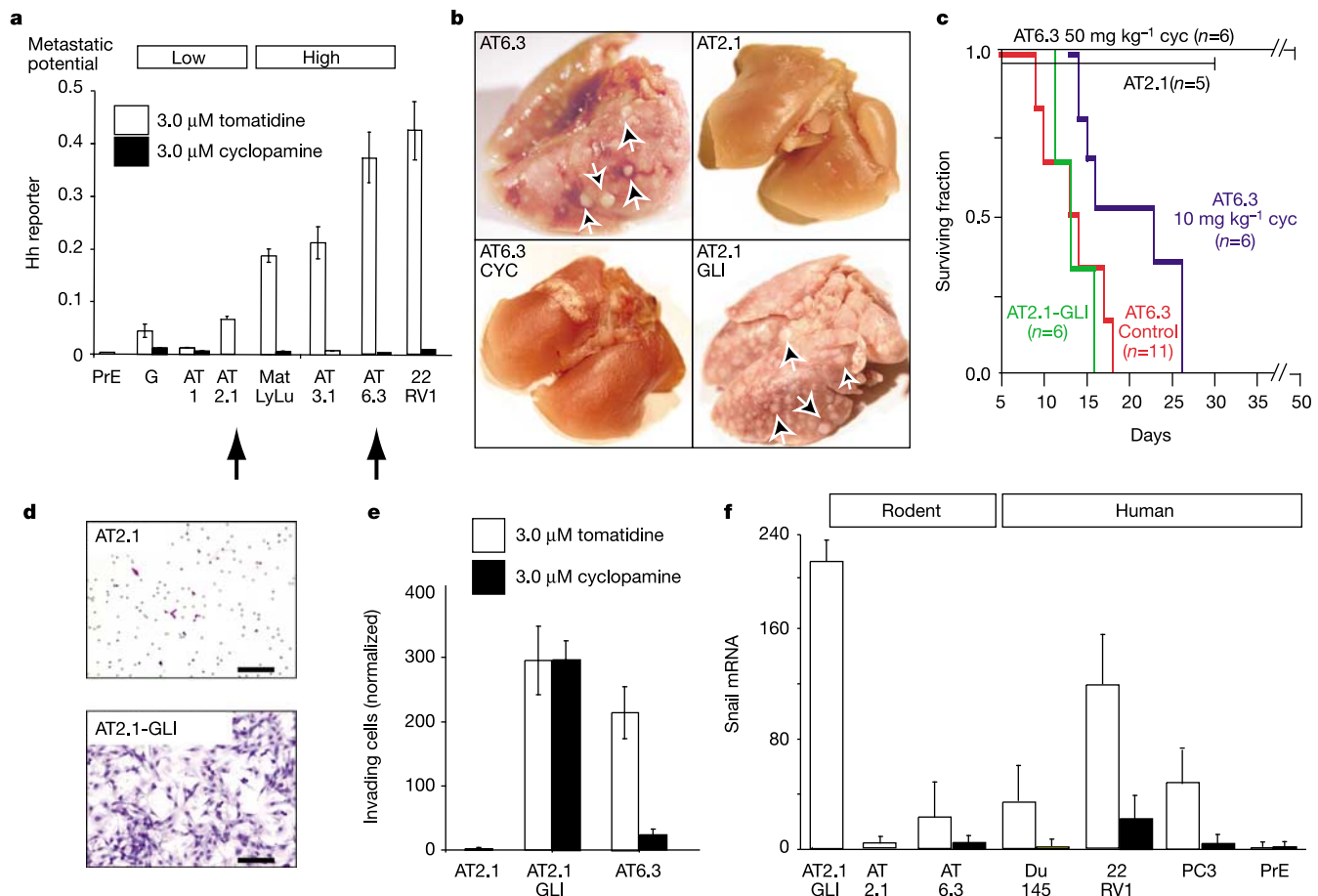


Figure 4 Hh pathway activity determines metastatic potential. **a**, Hh reporter activity in rodent prostate cancer lines with high (Mat-LyLu, AT3.1, AT6.3) and low (G, AT1, AT2) metastatic potential. Subsequent experiments used AT2.1 and AT6.3 (arrows). **b**, Lung metastases (arrows) in control mice subcutaneously inoculated with AT6.3 cells. Cyclopamine prevented lung metastasis (AT6.3 CYC). AT2.1 tumours did not metastasize, unless Hh pathway activity was augmented by GLI overexpression (AT2.1 GLI). **c**, Survival of nude mice bearing subcutaneous Dunning prostate carcinoma xenografts. AT6.3 tumours were treated intraperitoneally with vehicle alone or with 10 or 50 mg kg⁻¹ d⁻¹ cyclopamine. Median survival times in controls and animals treated with 10 mg kg⁻¹ were 13.5 and 22 d, respectively. No lethality was observed with the 50 mg kg⁻¹ dose. Untreated AT2.1 tumours were not lethal, whereas mice bearing

AT2.1-GLI tumours showed a median survival of 13 d. **d-f**, Hh pathway activation drives a metastasis-promoting program of cell invasiveness and gene expression. **d**, Photomicrographs of a modified Boyden chamber assay show numerous AT2.1-GLI cells (top), but few AT2.1 cells (bottom) that have traversed a Matrigel-coated membrane after 20 h. Scale bar, 100 μ M. **e**, Control-treated AT6.3 cells readily invaded the membrane, but cyclopamine treatment blocked invasion. The number of cells that have invaded is normalized to total viable cell mass on both sides of the membrane. **f**, Quantitative real-time RT-PCR showing decreased expression of *Snail* mRNA in AT6.3 cells treated with cyclopamine and enhanced expression in AT2.1-GLI. Error bars indicate s.e.m.

SMO expression may be a focal point of regulation in tissue regeneration and tumorigenesis. The dual roles of Hh pathway activity in promoting growth and metastasis of prostate cancer suggest that assessment and manipulation of Hh pathway activity may provide an important clinical avenue for the diagnosis and treatment of potentially lethal cancers.

Note added in proof: During the final revision of our manuscript, two new publications appeared, implicating Hh signalling in prostate cancer growth^{31,32}.

Methods

Cells and tissues

PC3, CWR22RV1, DU145 and LnCAP (ATCC) cells, were cultured in growth media (RPMI-1640 supplemented with 10% fetal bovine serum). AT6.3 and AT2.1 cells were cultured in growth media supplemented with 250 nM dexamethasone. PrE cells (Cambrex Biochemicals) were cultured according to the vendor's instructions. Human tissue samples are described in Supplementary Tables 1 and 2.

RNA isolation and analysis

Total cellular RNA was isolated and used to synthesize random primed first strand complementary DNA for analysis by conventional and quantitative real-time (SYBR green

stain) PCR (qRT-PCR) as described¹³. Amplification of Hh pathway components was normalized in qRT-PCR experiments to that of endogenous phosphoglycerate kinase in each sample. Oligonucleotide primers are listed in Supplementary Table 4.

Reporter assays

Gli-luciferase reporter assays were performed as described¹¹.

Stable transfections

Cells were transfected in 100-mm dishes with 15 μ l of Fugene6 transfection reagent (Roche) and 5 μ g of plasmid DNA, consisting of pKO-Neo (Invitrogen) alone or in a 1:19 ratio with either pSR α -FLAG-Gli1, pSR α -FLAG-Gli1ZFD (ref. 14), pGEM-mSMO-EAN or pIC-LacZ. Transfectants were selected with Geneticin (500 μ g ml⁻¹; Gibco) and subcloned.

Viability assays

Viable cell mass was assayed using the CellTiter96 reagent (Promega) as described⁸.

Xenografts

All animal studies were carried out using approved institutional protocols. CWR22RV1 ($n = 14$) and PC3 tumour xenografts ($n = 20$) were inoculated subcutaneously (s.c.) with 2.5×10^5 cells at two sites per female athymic mouse, treated and measured as described⁸. For Ki67 staining, groups of animals bearing 400–500-mm³ tumours (average volume) were injected s.c. with vehicle alone or with cyclopamine (10 mg kg⁻¹ d⁻¹) for 9 (PC3) or

10 d (CWR22RV1). In tumour regression studies, CWR22RV1 ($n = 20$), CWR22RV1GLI ($n = 8$) and PC3 ($n = 12$) tumours were grown to an average volume of 195 mm^3 and treated with $50 \text{ mg kg}^{-1} \text{ d}^{-1}$ cyclopamine or vehicle. Treatment was stopped after 28 d (PC3) or 22 d (22RV1), 7 d after all tumours appeared to have completely regressed. For metastasis studies, AT6.3, AT 2.1 and AT2.1-GLI rat prostate cancer cells in PBS were inoculated s.c. but without Matrigel, and treatment was started the next day with daily injections. Injections comprised intraperitoneal (i.p.) corn oil vehicle (Sigma) alone (AT2.1, $n = 5$; AT6.3, $n = 6$; and AT2.1-GLI, $n = 5$) or with cyclopamine ($10 \text{ mg kg}^{-1} \text{ d}^{-1}$ or $50 \text{ mg kg}^{-1} \text{ d}^{-1}$; AT6.3; $n = 12$ per dose), or s.c. cyclopamine at $50 \text{ mg kg}^{-1} \text{ d}^{-1}$ (AT6.3; $n = 5$).

Prostate regeneration

C57Bl6/J mice were castrated, rested for 7 d and treated with daily s.c. vehicle (80% glycerol trioleate in ethanol) alone, with DHT (50 mg kg^{-1}), with separate injections of DHT (s.c.) and 5E1mAb ($150 \mu\text{g d}^{-1}$ i.p.) or of DHT (s.c.) and cyclopamine (s.c.) (50 mg kg^{-1}) for 10 d. Ventral prostates were collected, weighed and processed for histology.

In vitro invasion assays

Cells were pre-treated with $3 \mu\text{M}$ cyclopamine or tomatidine for 24 h, and 2×10^5 cells were loaded into the top of a 24-well Matrigel invasion chamber assay plate (BD Biocoat). Cells reaching the lower chamber were counted and results were normalized to viable cell mass assayed as described above.

Ki-67 staining

Sections prepared from control- and cyclopamine-treated tumours were incubated with rabbit polyclonal antisera against Ki-67 (NovoCastra). Immunodetection was performed with the VectaStain ABC kit (Vector Laboratories). The ratio of Ki-67-positive to total nuclei was calculated in at least 300 cells examined in each of five randomly selected regions.

Received 1 February; accepted 23 August 2004; doi:10.1038/nature02962.

Published online 12 September 2004.

- Ingham, P. W. & McMahon, A. P. Hedgehog signaling in animal development: paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001).
- Taipale, J. & Beachy, P. A. The Hedgehog and Wnt signalling pathways in cancer. *Nature* **411**, 349–354 (2001).
- Litingtung, Y., Lei, L., Westphal, H. & Chiang, C. Sonic hedgehog is essential to foregut development. *Nature Genet.* **20**, 58–61 (1998).
- Lamm, M. L. *et al.* Sonic hedgehog activates mesenchymal Gli1 expression during prostate ductal bud formation. *Dev. Biol.* **249**, 349–366 (2002).
- Freestone, S. H. *et al.* Sonic hedgehog regulates prostatic growth and epithelial differentiation. *Dev. Biol.* **264**, 352–362 (2003).
- Berman, D. M. *et al.* Roles for Hedgehog signaling in androgen production and prostate ductal morphogenesis. *Dev. Biol.* **267**, 387–398 (2004).
- Lum, L. & Beachy, P. A. The Hedgehog response network: sensors, switches, and routers. *Science* **304**, 1755–1759 (2004).
- Berman, D. M. *et al.* Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. *Nature* **425**, 846–851 (2003).
- Thayer, S. P. *et al.* Hedgehog is an early and late mediator of pancreatic cancer tumorigenesis. *Nature* **425**, 851–856 (2003).
- Watkins, D. N. *et al.* Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature* **422**, 313–317 (2003).
- Taipale, J. *et al.* Effects of oncogenic mutations in Smoothened and Patched can be reversed by cyclopamine. *Nature* **406**, 1005–1009 (2000).
- Chen, J. K., Taipale, J., Cooper, M. K. & Beachy, P. A. Inhibition of Hedgehog signaling by direct binding of cyclopamine to Smoothened. *Genes Dev.* **16**, 2743–2748 (2002).
- Berman, D. M. *et al.* Medulloblastoma growth inhibition by hedgehog pathway blockade. *Science* **297**, 1559–1561 (2002).
- Park, H. L. *et al.* Mouse Gli1 mutants are viable but have defects in SHH signaling in combination with a Gli2 mutation. *Development* **127**, 1593–1605 (2000).
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J. & Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. *Proc. Natl Acad. Sci. USA* **100**, 3983–3988 (2003).
- Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
- Kachinsky, A. M., Dominov, J. A. & Miller, J. B. Myogenesis and the intermediate filament protein, nestin. *Dev. Biol.* **165**, 216–228 (1994).
- Lendahl, U., Zimmerman, L. B. & McKay, R. D. CNS stem cells express a new class of intermediate filament protein. *Cell* **60**, 585–595 (1990).
- Zulewski, H. *et al.* Multipotential nestin-positive stem cells isolated from adult pancreatic islets differentiate *ex vivo* into pancreatic endocrine, exocrine, and hepatic phenotypes. *Diabetes* **50**, 521–533 (2001).
- Molofsky, A. V. *et al.* Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962–967 (2003).
- English, H. F., Santen, R. J. & Isaacs, J. T. Response of glandular versus basal rat ventral prostatic epithelial cells to androgen withdrawal and replacement. *Prostate* **11**, 229–242 (1987).
- Meeker, A. K., Sommerfeld, H. J. & Coffey, D. S. Telomerase is activated in the prostate and seminal vesicles of the castrated rat. *Endocrinology* **137**, 5743–5746 (1996).
- Garraway, L. A. *et al.* Intermediate basal cells of the prostate: *in vitro* and *in vivo* characterization. *Prostate* **55**, 206–218 (2003).
- Isaacs, J. T., Isaacs, W. B., Feitz, W. F. & Scheres, J. Establishment and characterization of seven Dunning rat prostatic cancer cell lines and their use in developing methods for predicting metastatic abilities of prostatic cancers. *Prostate* **9**, 261–281 (1986).
- Dong, J. T. *et al.* KAI1, a metastasis suppressor gene for prostate cancer on human chromosome 11p11.2. *Science* **268**, 884–886 (1995).

- Albini, A. *et al.* A rapid *in vitro* assay for quantitating the invasive potential of tumor cells. *Cancer Res.* **47**, 3239–3245 (1987).
- Guan, R. J. *et al.* Drg-1 as a differentiation-related, putative metastatic suppressor gene in human colon cancer. *Cancer Res.* **60**, 749–755 (2000).
- Cano, A. *et al.* The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nature Cell Biol.* **2**, 76–83 (2000).
- Bandyopadhyay, S. *et al.* The Drg-1 gene suppresses tumor metastasis in prostate cancer. *Cancer Res.* **63**, 1731–1736 (2003).
- Bernards, R. & Weinberg, R. A. A progression puzzle. *Nature* **418**, 823 (2002).
- Fan, L. *et al.* Hedgehog signaling promotes prostate xenograft tumor growth. *Endocrinology* **145**, 3961–3970 (2004).
- Sanchez, P. *et al.* Inhibition of prostate cancer proliferation by interference with SONIC HEDGEHOG-GLI1 signaling. *Proc. Natl Acad. Sci. USA* **101**, 12561–12566 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank W. Bushman for sharing results before publication, S. Dalrymple, H. Fedor, M. Gerstenblith, T. Harcke, W. Kleeberger, R. Montes de Oca, E. Traband, Y. Xin and K. Young for help with experiments; M. Eisenberger, M. Carducci and W. Isaacs for support of autopsy studies; C. Sawyers, W. Nelson and D. Neil Watkins for discussions and A. Joyner, D. Coffey and J. Chen for reagents. We are very grateful to A. De Marzo and the Johns Hopkins Prostate Specimen Repository for generously providing tissue samples. This research was supported by NIH and the Prostate Cancer Foundation. P.A.B. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to D.M.B. (dberman@jhmi.edu) or P.A.B. (pbeachy@jhmi.edu).

p19^{ARF} directly and differentially controls the functions of c-Myc independently of p53

Ying Qi, Mark A. Gregory, Zhaoliang Li, Jeffrey P. Brousal, Kimberly West & Stephen R. Hann

Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-2175, USA

Increased expression of the oncogenic transcription factor c-Myc causes unregulated cell cycle progression¹. c-Myc can also cause apoptosis, but it is not known whether the activation and/or repression of c-Myc target genes mediates these diverse functions of c-Myc. Because unchecked cell cycle progression leads to hyperproliferation and tumorigenesis, it is essential for tumour suppressors, such as p53 and p19^{ARF} (ARF), to curb cell cycle progression in response to increased c-Myc (refs 2, 3). Increased c-Myc has previously been shown to induce ARF expression, which leads to cell cycle arrest or apoptosis through the activation of p53 (ref. 4). Here we show that ARF can inhibit c-Myc by a unique and direct mechanism that is independent of p53. When c-Myc increases, ARF binds with c-Myc and dramatically blocks c-Myc's ability to activate transcription and induce hyperproliferation and transformation. In contrast, c-Myc's ability to repress transcription is unaffected by ARF and c-Myc-mediated apoptosis is enhanced. These differential effects of ARF on c-Myc function suggest that separate molecular mechanisms mediate c-Myc-induced hyperproliferation and apoptosis. This direct feedback mechanism represents a p53-independent checkpoint to prevent c-Myc-mediated tumorigenesis.

When ARF was examined as a nucleolar marker during studies examining c-Myc localization in immortalized mouse embryo fibroblasts (MEF) lacking p53, we observed that ARF shifted localization in response to increased c-Myc. In untransfected cells, endogenous ARF protein was found in nucleoli and colocalized with

nucleolin (Fig. 1a–c), as has previously been shown⁵. In contrast, in cells expressing increased levels of c-Myc-YFP (yellow fluorescent protein), ARF colocalized with c-Myc-YFP in a diffuse nucleoplasmic staining pattern (Fig. 1d–f). Higher expression of c-Myc resulted in a more complete exclusion of ARF from nucleoli

(Fig. 1e, lower cell). Transfection of untagged c-Myc demonstrated that the localization shift is not a result of the fluorescent tag (Fig. 1g–i). In addition, the shift of endogenous ARF also occurred in wild-type MEF cells (Fig. 1j–l), suggesting that the p53 status of the cells has no effect on ARF relocation caused by c-Myc.

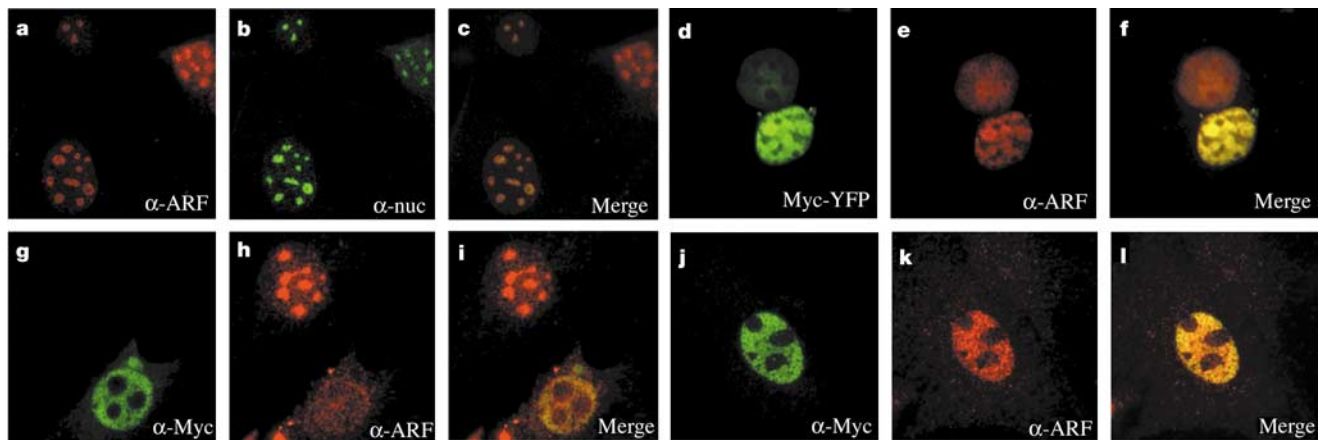


Figure 1 Endogenous ARF colocalizes with c-Myc in the nucleoplasm upon c-Myc overexpression. **a–c**, $p53^{-/-}$ MEFs immunostained for ARF and nucleolin. **d–f**, $p53^{-/-}$ MEFs transfected with c-Myc-YFP and immunostained for ARF. **g–i**, $p53^{-/-}$ MEFs

transfected with c-Myc and immunostained for c-Myc and ARF. **j–l**, Wild-type MEFs transfected with c-Myc and immunostained for c-Myc and ARF. Cells were visualized by fluorescence microscopy.

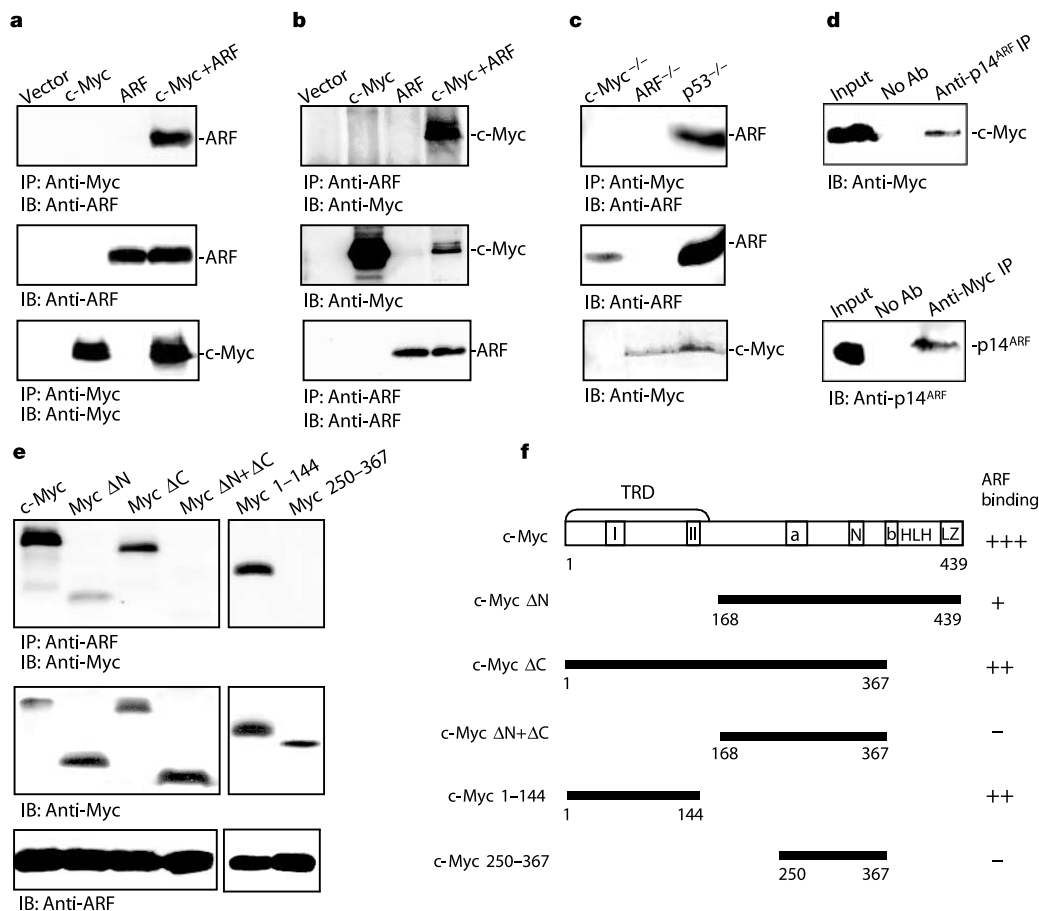


Figure 2 ARF binds to c-Myc. **a–d**, Coimmunoprecipitation of ARF and c-Myc. Cos-7 cells transfected with c-Myc and/or ARF were subjected to immunoprecipitation (IP) using anti-Mycfl (fl, full length; **a**) or anti-ARF (**b**) followed by immunoblot (IB) analysis using the indicated antibody. **c**, $c-myc^{-/-}$ cells, $ARF^{-/-}$ MEFs or $p53^{-/-}$ MEFs were subjected to IP using anti-Mycfl followed by IB analysis. **d**, HeLa cells were subjected to IP using

anti-p14^{ARF} or anti-Mycfl followed by IB analysis. **e**, Coimmunoprecipitation of c-Myc deletions with ARF. Cos-7 cells transfected with the indicated c-Myc deletion mutant were subjected to IP with anti-ARF followed by IB analysis. **f**, Schematic representation of the structure of c-Myc and c-Myc deletions showing relative binding to ARF.

Together, these results suggest that c-Myc interacts with ARF in the nucleoplasm and prevents its nucleolar localization. However, in other cell lines we have also observed colocalization of ARF and c-Myc in nucleoli on ectopic expression of ARF (M.A.G. and S.R.H., unpublished observations). Datta *et al.* have also observed colocalization of c-Myc and ARF in nucleoli⁶. The mechanism and role for this differential localization of c-Myc and ARF are under investigation.

To determine whether the two proteins interact, we expressed c-Myc and ARF individually or together and immunoprecipitation was performed using c-Myc antibody under stringent detergent conditions. Immunoblot analysis was then performed using ARF antibody. When coexpressed, ARF coimmunoprecipitated with c-Myc (Fig. 2a). In the converse experiment, using ARF antibody for immunoprecipitation, c-Myc coprecipitated when coexpressed with ARF (Fig. 2b). Using *ARF*^{-/-} and *c-myc*^{-/-} cells as negative controls, we also demonstrated that endogenous ARF coimmunoprecipitates with endogenous c-Myc in *p53*^{-/-} MEFs (Fig. 2c). Endogenous human p14^{ARF} also specifically coimmunoprecipitated with endogenous human c-Myc from HeLa cells (Fig. 2d).

To identify the domains of c-Myc that interact with ARF, coimmunoprecipitation assays were carried out with a panel of c-Myc proteins having deletions, including deletions of the C-terminal helix-loop-helix/leucine zipper (HLH/LZ) domain, which is necessary for heterodimerization with c-Myc's partner Max, and the amino-terminal transcriptional regulatory domain (TRD), which is critical for transcriptional activation and repression⁷. The c-Myc proteins with deletions of the TRD (ΔN) or the HLH/LZ (ΔC) domains were still able to bind ARF, albeit less efficiently

(Fig. 2e, left panel), as did all other c-Myc proteins containing different deletions throughout the protein (data not shown). However, deletion of both the TRD and HLH/LZ ($\Delta N + \Delta C$) resulted in no detectable binding (Fig. 2e, left panel). In addition, an N-terminal fragment (amino acids 1–144) representing the TRD bound efficiently to ARF; whereas an internal fragment (amino acids 250–367) containing the acidic domain showed no detectable binding (Fig. 2e, right panel). Because deletion of the TRD had a greater impact on ARF binding (Fig. 2f), the TRD may represent the primary binding site. These experiments demonstrate that both exogenous and endogenous c-Myc and ARF associate in a highly stable complex and that ARF binds to both the TRD and HLH/LZ domains of c-Myc.

ARF binds to regions of c-Myc that are critical for its transcriptional activity, so we next examined the effect of ARF on the activity of c-Myc responsive promoters. In MEF cells lacking ARF (*ARF*^{-/-}), c-Myc alone increased the activation of the telomerase reverse transcriptase (*htert*) promoter by approximately threefold (Fig. 3a), which is consistent with earlier reports^{8,9}. Coexpression of ARF completely blocked the ability of c-Myc to transactivate (Fig. 3a). When c-Myc was expressed in MEF cells having elevated endogenous ARF (*p53*^{-/-}), c-Myc also failed to activate the *htert* promoter (Fig. 3a). Coexpression of exogenous ARF in these cells resulted in further inhibition of *htert* promoter activity (Fig. 3a). ARF also blocked the ability of c-Myc to transactivate another c-Myc-responsive promoter, *cull1*¹⁰ (Supplementary Fig. 1a), confirming that this effect of ARF is not promoter-specific. In addition, as endogenous ARF levels increase during passaging of primary MEFs⁴, there is a concordant decrease in c-Myc transactivation.

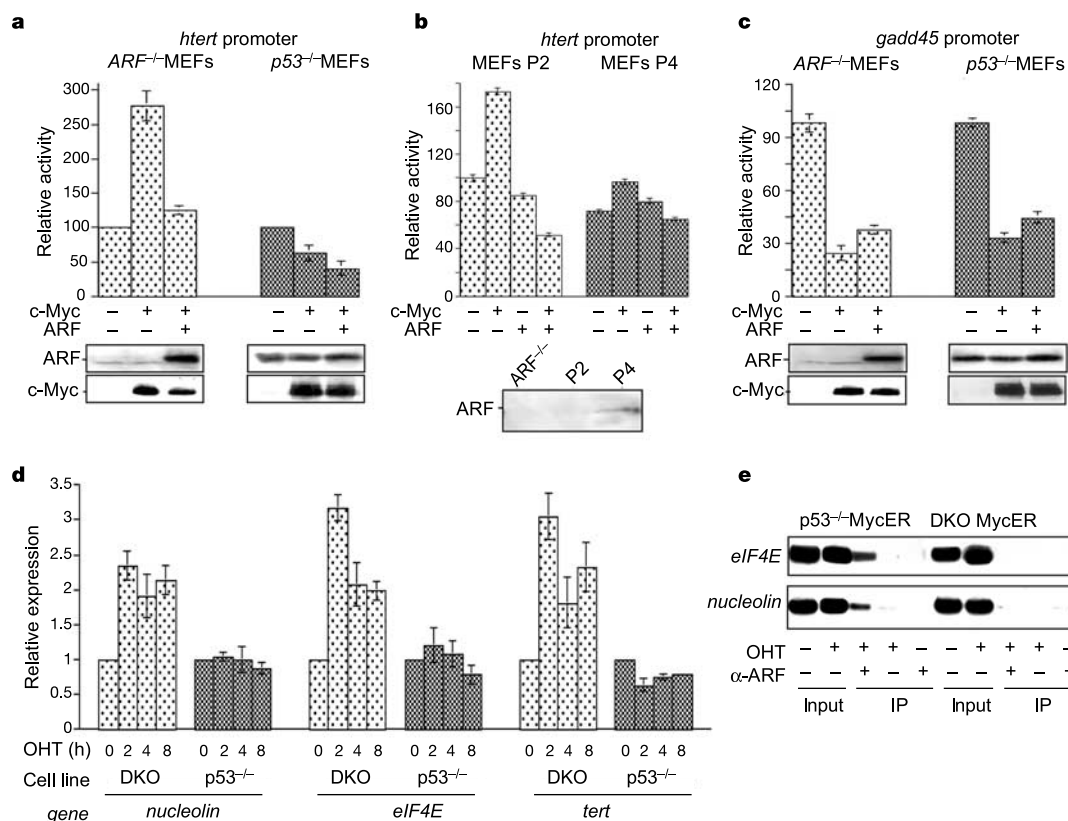


Figure 3 ARF differentially regulates the transcriptional activities of c-Myc. Reporter constructs for *htert* (a) and *gadd45* (b) promoters were transfected into *ARF*^{-/-} MEFs, *p53*^{-/-} MEFs or embryonic MEFs at the indicated passage number (b), with or without c-Myc or ARF, and reporter activity was determined. d, Quantitative real-time PCR analysis of c-Myc target gene expression in *p53*^{-/-} MycER MEFs and DKO

MycER MEFs with or without c-MycER activation ($\pm$ OHT). ARF and c-MycER expression in these cell lines was confirmed by IB analysis (Supplementary Fig. 3a and b). e, Recruitment of ARF to c-Myc target gene promoters. Chromatin prepared from *p53*^{-/-} MycER and DKO MycER cells ($\pm$ OHT) was subjected to IP using anti-ARF followed by PCR using primers for the *eIF4E* and *nucleolin* promoters.

c-Myc transactivated the *htert* promoter in early passage MEF cells (P2); however, in later passage MEF cells (P4) with higher levels of ARF protein, c-Myc was unable to activate (Fig. 3b). In contrast, there was no effect of exogenous or endogenous ARF on repression of the *gadd45*¹¹ promoter by c-Myc (Fig. 3c). We obtained the same results using another c-Myc-responsive promoter, *pdgfr β* ¹² (Supplementary Fig. 1a). These results suggest that ARF effectively blocks the transactivation function of c-Myc, but not c-Myc's ability to repress transcription.

We next wanted to determine the effect of ARF on the regulation of endogenous c-Myc target genes in a defined p53 null background. Thus, we compared p53/ARF double knockout (DKO) and p53^{-/-} MEFs that express the chimaeric c-MycER protein (ER, oestrogen receptor), whose activity is inducible upon hydroxytamoxifen (OHT) treatment. Immunofluorescence analysis demonstrated that ARF relocates to the nucleoplasm upon activation of c-MycER (data not shown), as shown for transiently expressed c-Myc (Fig. 1). Real-time polymerase chain reaction (PCR) analyses revealed that activated c-MycER induced the expression of *nucleolin*, *elF4E*, *tert*, *cdk4* and *cull1* (Fig. 3d and Supplementary Fig. 1b) in DKO cells, but there was no upregulation of these genes in p53^{-/-} MEF cells, which have high levels of endogenous ARF. In contrast, the downregulation of *gadd45* and *p15^{INK4b}* by activated c-MycER was not significantly different in DKO compared to p53^{-/-} MEFs (Supplementary Fig. 1b). In addition, northern blot analyses of Rat1a cells demonstrated that exogenous ARF also blocks upregulation of c-Myc target genes, *elF4E* and *cad*, by activated c-MycER, but does not affect repression of the *gadd45* gene (Supplementary Fig. 1c).

ARF effectively blocked the induction of c-Myc target genes and colocalizes with c-Myc in the nucleoplasm, so we wanted to determine whether ARF protein associates with c-Myc protein at the promoters of c-Myc target genes. Chromatin immunoprecipitation (ChIP) analysis revealed that ARF protein specifically associates with the promoters of *elF4E* and *nucleolin* upon activation of c-MycER in p53^{-/-} MEFs (Fig. 3e). We found no difference between p53^{-/-} and DKO MEFs in ChIP assays using c-Myc antibody (data not shown), suggesting that the presence of ARF does not influence the recruitment of c-Myc to target promoters. c-Myc only binds to target promoters as a heterodimer with Max, so these results suggest that ARF forms a complex with c-Myc and Max. To confirm this, we performed coimmunoprecipitation experiments with Max antibody and found that ARF coprecipitated with c-Myc and Max (Supplementary Fig. 2). Taken together, these results illustrate that the interaction of ARF with c-Myc differentially controls c-Myc activities by blocking the induction of c-Myc target genes at the promoter without affecting the repression of c-Myc target genes.

We next wanted to determine whether ARF influences c-Myc biological functions, independently of p53. We examined the proliferation rate of the DKO and p53^{-/-} MEFs expressing c-MycER. Without activation of c-MycER, the p53^{-/-} and DKO MEFs proliferate at similar rates (Fig. 4a). However, upon activation of c-MycER, the p53^{-/-} MEFs accumulated at a slower rate than the control cells, whereas the DKO cells proliferated at a much higher rate (Fig. 4a). There was also a significant increase in the number of apoptotic cells over time upon activation of c-MycER in p53^{-/-} cells, whereas there was no accumulation of apoptotic cells in the

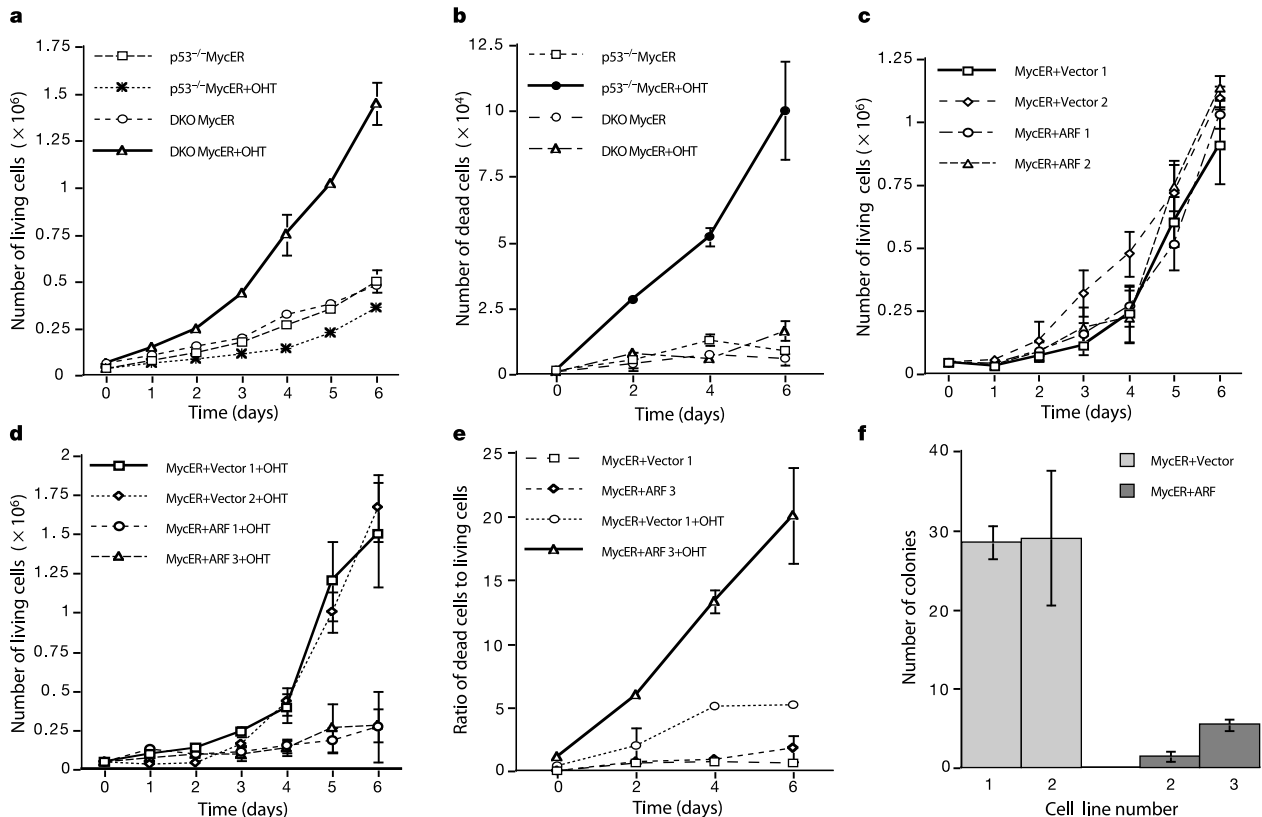


Figure 4 ARF inhibits hyperproliferation and transformation by c-Myc, yet facilitates c-Myc-induced apoptosis. **a**, Proliferation assay of p53^{-/-} MycER and DKO MycER MEFs with or without c-MycER activation ($\pm$ OHT). **b**, p53^{-/-} MycER or DKO MycER MEFs ($\pm$ OHT) were assayed for apoptotic cell death. **c**, Proliferation assay of Rat1a MycER $\pm$ ARF cell lines without MycER activation. ARF and c-MycER expression in these

cell lines was confirmed by IB analysis (Supplementary Fig. 3a and b). **d**, Proliferation assay of Rat1a cell lines with MycER activation. **e**, Rat1a cell lines ($\pm$ OHT) were assayed for apoptosis in low serum. **f**, Rat1a cell lines (+OHT) were plated in soft agar and analysed for colony growth.

DKO or control MEFs (Fig. 4b). These results suggest that ARF inhibits c-Myc-induced hyperproliferation of MEF cells, while enhancing c-Myc-induced apoptosis, independently of p53.

The effects of ARF on c-Myc functions were also examined in Rat1a cells that have been used by us¹³ and others¹⁴ to demonstrate that c-Myc causes hyperproliferation, transformation and apoptosis in low serum. These cells express low levels of endogenous ARF (Supplementary Fig. 3a), but do not induce the expression of p53 or p21^{CIP1} proteins in response to c-MycER activation or ARF overexpression (Supplementary Fig. 3c). Although the oncogenic pathway of p53 activation appears to be disabled in Rat1a cells, DNA damage from doxorubicin treatment induces low levels of p53 and p21^{CIP1} (Supplementary Fig. 3c). These results are in agreement with previous reports that found c-Myc-mediated apoptosis in Rat1a cells to be p53-independent^{15,16}. Overexpression of ARF had no effect on Rat1a cellular proliferation in the absence of c-MycER activation (Fig. 4c). However, ARF coexpression effectively blocked activated c-MycER-induced hyperproliferation (Fig. 4d). When the Rat1a cells were shifted to media containing low serum at the time c-Myc was activated, the number of apoptotic cells relative to living cells increased dramatically over time (Fig. 4e). As with the MEF cells, these results probably reflect the combined effects of ARF enhancing c-Myc-induced apoptosis while inhibiting c-Myc-induced proliferation of the remaining living cells. Without c-Myc activation, ARF overexpression had no effect on apoptosis (Fig. 4e). In addition, coexpression of ARF sharply reduced the ability of activated c-MycER to induce transformation, as measured by anchorage-independent growth in soft agar (Fig. 4f). These results further confirm that ARF differentially controls c-Myc biological activities, independently of p53.

Although p53 has been proposed to mediate c-Myc-induced apoptosis and ARF-induced apoptosis and growth arrest⁴, there are numerous reports demonstrating p53-independent apoptosis *in vivo* and *in vitro* caused by c-Myc or ARF^{15–23}. Our results suggest that when ARF is induced by elevated c-Myc, it prevents hyperproliferation and transformation through a direct negative-feedback mechanism. Therefore, the loss of ARF could contribute to c-Myc-induced tumorigenesis by at least two mechanisms. Consistent with this idea, the downregulation of ARF/INK4a by Bmi-1 is probably responsible for the strong collaboration of the *bmi-1* and *c-myc* oncogenes to induce murine lymphomagenesis²⁴. In addition, *Em-myc*-induced lymphomagenesis is greatly accelerated by ARF loss²⁵, and inactivation of ARF was consistently found in murine myeloid tumours arising from deregulated c-Myc²⁶.

Our results suggest that ARF blocks c-Myc-induced hyperproliferation and transformation by blocking transactivation of key c-Myc target genes. In contrast, the inability of ARF to block c-Myc-induced apoptosis suggests that c-Myc-induced apoptosis is mediated through another mechanism, perhaps transrepression of specific anti-apoptotic genes. Previous reports also suggest that the mechanism for c-Myc-induced apoptosis is distinct from the mechanism of c-Myc-stimulated proliferation, and that this mechanism is dependent on c-Myc transrepression^{13,14,27–29}. Therefore, in addition to ARF mediating p53 activation, ARF binding to c-Myc represents an important fail-safe mechanism for preventing aberrant c-Myc signalling and tumorigenesis through differential control of c-Myc transcriptional activities. □

Methods

Details regarding expression constructs, cell culture, transfection/infection conditions, and growth and apoptosis assays can be found in the Supplementary Information.

Immunoprecipitation and immunoblotting analysis

Cell lysates were prepared and coimmunoprecipitation of endogenous proteins was performed as described previously⁷. For coimmunoprecipitation of exogenous proteins, cells were lysed in Ab lysis buffer (20 mM Tris, pH 7.5, 0.5% Triton X-100, 0.5% deoxycholic acid (DOC) and 0.5% SDS, 1 mM EDTA) and immunoprecipitates were washed in RIPA buffer³⁰. The proteins were resolved by 15% SDS-PAGE and subjected to

immunoblot analysis using anti-c-Mycfl (06-340; Upstate) or anti-ARF (Ab80; GeneTex or 07-543; Upstate) and enhanced chemiluminescence for detection.

Immunofluorescence microscopy

The indicated cells were grown on glass coverslips, fixed and permeabilized as described previously⁷. Cells were incubated with anti-nucleolin (C23, MS-3; Santa Cruz), anti-ARF (Ab80) or anti-c-Myc (C-33; Santa Cruz) at a dilution of 1 µg ml⁻¹, and then incubated with the appropriate fluorescence-labelled secondary antibodies, AlexaFluor488 goat anti-mouse IgG or AlexaFluor594 donkey anti-rabbit IgG (Molecular Probes), at a dilution of 1:1000. Fluorescence microscopy was performed as described previously⁵ using a 63× objective.

Reporter assays

For luciferase assays, ARF^{-/-} MEFs or p53^{-/-} MEFs were seeded at 1 × 10⁵ cells per 35 mm dish. The next day, cells were transfected with 1.2 µg of c-Myc and/or 0.3 µg ARF expression vector and 1.2 µg of reporter construct, pRL-TK (500 ng) was included as an internal control. Luciferase assays were carried out according to the manufacturer's instructions (Dual-Luciferase Reporter Assay System; Promega). Results were normalized for expression of pRL-TK as measured by Renilla luciferase activity. For SEAP (secreted placental alkaline phosphatase) assays, culture media was collected from cells 36–60 h after transfection, heated at 65 °C for 30 min, and clarified by centrifugation. Culture media was added to 2 × SEAP buffer (2 M diethanolamine, 1 mM MgCl₂, and 20 mM L-homoarginine) containing 57 mM p-nitrophenyl phosphate (Sigma104 phosphatase substrate) and SEAP activity was measured by spectrophotometry at wavelength 405 nm. Luciferase or SEAP activity from cells transfected with reporter gene alone was standardized to 100%. Normalized values from duplicate samples were reported as the mean ± s.d. Each assay is representative of at least three independent experiments.

Anchorage-independent growth assay

Rat1a MycER or Rat1a MycER + ARF cell lines were plated at 2 × 10⁴ cells per 35-mm dish in soft agar containing DMEM plus 10% FCS with 2 µM OHT. Colonies (76 µm or larger) from triplicate plates were counted using an Omnicon colony counter (Bausch & Lomb) on day 10 after plating. Data from the two different monoclonal cell lines shown is representative of at least four different cell lines.

Quantitative real-time PCR

Four µg of total RNA isolated from p53^{-/-} MycER MEFs and DKO MycER MEFs treated with 1 µM OHT for the indicated time was reverse-transcribed using the Access RT-PCR system (Promega). Quantitative real-time PCR was performed using the iCycler and SYBR Green dye (BioRad). Relative measurement of gene expression was calculated following manufacturer's instructions using the standard curve method. The specific primer sequences used are listed in Supplementary Information. Relative values compared to the unactivated control samples were graphed as the mean ± s.d. from triplicate assays. Each analysis was representative of at least two different monoclonal cell lines.

Chromatin immunoprecipitation (ChIP)

p53^{-/-} MycER MEFs and DKO MycER MEFs were treated with 5 µM OHT for 4 h as indicated and the cells were cross-linked with 1% formaldehyde at 37 °C for 10 min. ChIP assays were performed using anti-ARF (Ab80) and the ChIP assay kit (Upstate) according to the manufacturer's instructions. Immunoprecipitated DNA was purified using a QIAquick Spin Kit (Qiagen) and subjected to PCR amplification using specific primer sets for *elF4E* and *nucleolin* (listed in Supplementary Information).

Received 3 June; accepted 23 August 2004; doi:10.1038/nature02958.

Published online 8 September 2004.

- Oster, S. K., Ho, C. S., Soucie, E. L. & Penn, L. Z. The myc oncogene: Marvelously complex. *Adv. Cancer Res.* **84**, 81–154 (2002).
- Sherr, C. J. The INK4a/ARF network in tumour suppression. *Nature Rev. Mol. Cell Biol.* **2**, 731–737 (2001).
- Pelengaris, S., Khan, M. & Evan, G. c-MYC: More than just a matter of life and death. *Nature Rev. Cancer* **2**, 764–776 (2002).
- Zindy, F. *et al.* Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization. *Genes Dev.* **12**, 2424–2433 (1998).
- Weber, J. D., Taylor, L. J., Roussel, M. F., Sherr, C. J. & Bar-Sagi, D. Nucleolar Arf sequesters Mdm2 and activates p53. *Nature Cell Biol.* **1**, 20–26 (1999).
- Datta, A. *et al.* Myc-ARF interaction inhibits the functions of Myc. *J. Biol. Chem.* **279**, 36698–36707 (2004).
- Grandori, C. & Eisenman, R. N. Myc target genes. *Trends Biochem. Sci.* **22**, 177–181 (1997).
- Greenberg, R. A. *et al.* Telomerase reverse transcriptase gene is a direct target of c-Myc but is not functionally equivalent in cellular transformation. *Oncogene* **18**, 1219–1226 (1999).
- Wu, K. J. *et al.* Direct activation of TERT transcription by c-MYC. *Nature Genet.* **21**, 220–224 (1999).
- O'Hagan, R. C. *et al.* Myc-enhanced expression of Cull1 promotes ubiquitin-dependent proteolysis and cell cycle progression. *Genes Dev.* **14**, 2185–2191 (2000).
- Marhin, W. W., Chen, S., Facchini, L. M., Fornace, A. J. Jr & Penn, L. Z. Myc represses the growth arrest gene *gadd45*. *Oncogene* **14**, 2825–2834 (1997).
- Oster, S. K. *et al.* Myc is an essential negative regulator of platelet-derived growth factor beta receptor expression. *Mol. Cell Biol.* **20**, 6768–6778 (2000).
- Xiao, Q. *et al.* Transactivation-defective c-MycS retains the ability to regulate proliferation and apoptosis. *Genes Dev.* **12**, 3803–3808 (1998).
- Evan, G. I. *et al.* Induction of apoptosis in fibroblasts by c-myc protein. *Cell* **69**, 119–128 (1992).
- Harrington, E. A., Bennett, M. R., Fanidi, A. & Evan, G. I. c-Myc-induced apoptosis in fibroblasts is inhibited by specific cytokines. *EMBO J.* **13**, 3286–3295 (1994).
- Lenahan, M. K. & Ozer, H. L. Induction of c-myc mediated apoptosis in SV40-transformed rat fibroblasts. *Oncogene* **12**, 1847–1854 (1996).

17. Amanullah, A., Liebermann, D. A. & Hoffman, B. p53-independent apoptosis associated with c-Myc-mediated block in myeloid cell differentiation. *Oncogene* **19**, 2967–2977 (2000).
18. Fukasawa, K., Wiener, F., Vande Woude, G. F. & Mai, S. Genomic instability and apoptosis are frequent in p53 deficient young mice. *Oncogene* **15**, 1295–1302 (1997).
19. Korgaonkar, C., Zhao, L., Modestou, M. & Quelle, D. E. ARF function does not require p53 stabilization or Mdm2 relocalization. *Mol. Cell. Biol.* **22**, 196–206 (2002).
20. Sakamuro, D. *et al.* c-Myc induces apoptosis in epithelial cells by both p53-dependent and p53-independent mechanisms. *Oncogene* **11**, 2411–2418 (1995).
21. Trudel, M. *et al.* c-myc-induced apoptosis in polycystic kidney disease is Bcl-2 and p53 independent. *J. Exp. Med.* **186**, 1873–1884 (1997).
22. Tsuji, K. *et al.* p53-independent apoptosis is induced by the p19ARF tumor suppressor. *Biochem. Biophys. Res. Commun.* **295**, 621–629 (2002).
23. Weber, J. D. *et al.* p53-independent functions of the p19(ARF) tumor suppressor. *Genes Dev.* **14**, 2358–2365 (2000).
24. Jacobs, J. J. *et al.* Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. *Genes Dev.* **13**, 2678–2690 (1999).
25. Eischen, C. M., Weber, J. D., Roussel, M. F., Sherr, C. J. & Cleveland, J. L. Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis. *Genes Dev.* **13**, 2658–2669 (1999).
26. Haviernik, P., Schmidt, M., Hu, X. & Wolff, L. Consistent inactivation of p19(Arf) but not p15(Ink4b) in murine myeloid cells transformed *in vivo* by deregulated c-Myc. *Oncogene* **22**, 1600–1610 (2003).
27. Oster, S. K., Mao, D. Y., Kennedy, J. & Penn, L. Z. Functional analysis of the N-terminal domain of the Myc oncoprotein. *Oncogene* **22**, 1998–2010 (2003).
28. Conzen, S. D. *et al.* Induction of cell cycle progression and acceleration of apoptosis are two separable functions of c-Myc: transrepression correlates with acceleration of apoptosis. *Mol. Cell. Biol.* **20**, 6008–6018 (2000).
29. Soucek, L. *et al.* Omomyc, a potential Myc dominant negative, enhances Myc-induced apoptosis. *Cancer Res.* **62**, 3507–3510 (2002).
30. Gregory, M. A., Qi, Y. & Hann, S. R. Phosphorylation by glycogen synthase kinase-3 controls c-myc proteolysis and subnuclear localization. *J. Biol. Chem.* **278**, 51606–51612 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank C. Sherr and S. Hiebert for pCMV5-ARF and ARF^{-/-} MEFs, G. Zambetti for DKO MEFs, R. DePinho for *htert*-SEAP and *cull1*-luc, L. Penn for *gadd45*-luc and *pdgfr*-luc, E. Raley for p53^{-/-} MEFs, J. Pieterpol for REF 112 cells, and J. Sedivy for HO16 cells. We also thank R. Eisenman for critical review of the manuscript. This work was supported by grants from NIH to S.R.H.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.R.H. (steve.hann@vanderbilt.edu).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

Natureevents: Silke Opstrup (4994)

France/ Switzerland:

Amelie Pequinot (4974)

Production Manager:

Billie Franklin
To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Stefan Hales

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

The philanthropic principle

Government and industry create the majority of scientific jobs. But over the past few years, non-profit research institutions, funded by philanthropists, have boosted employment — especially in the United States.

There is, for example, the Stowers Institute for Medical Research in Kansas City, Missouri, which only opened its doors in 2000, but which this year announced plans to house another 20 research groups — doubling its capacity (see overleaf). And the high-profile Howard Hughes Medical Institute (HHMI) this week began recruiting 24 group leaders for its forthcoming Janelia Farm Research Campus in northern Virginia.

Both of these sites are interesting, not just for their billions of dollars in endowments, but for the strikingly different career paths they create. The Stowers provides start-up funds, but encourages its investigators to win grants; it also has a tenure-track system in place and has recruited gradually. The HHMI-funded Janelia scientists, by contrast, may be forbidden to compete for US National Institutes of Health grants, will be given fixed-term, renewable appointments and will mostly be signed up before Janelia opens in 2006.

That is one advantage of foundation-funded research institutions: they can try different models, and adjust as necessary. They are also less susceptible to political and economic pressure. For example, Stowers scientists can pursue stem-cell research, even though the Missouri state government will not fund such activities. And scientists at both campuses are less dependent on federal funds, and so less affected by small dips in the economy — although the founding trusts still need a fairly stable stock market to thrive.

With philanthropic ventures such as the Bill and Melinda Gates Foundation becoming increasingly interested in science funding, the outlook for jobs in foundation-funded centres seems bright.

Paul Smaglik
Naturejobs editor



Contents

REGIONS

Missouri's merits p720

CAREER VIEW

Nuts & Bolts

Totally stumped?

Graduate Journal

The science of religion

Movers

Nouria Hernandez and
Winship Herr p722

WWW.NATUREJOBS.COM

Career centre

Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Centre stage in Missouri

St Louis considers itself the gateway to the west; Kansas City, the geographic heart of the United States. But leaders in both Missouri cities would rather the state were known as a premier centre for research. Two private initiatives in the past few years have pledged more than \$3 billion to help them rival their competition on the coasts. The funds, for Washington University in St Louis and the Stowers Institute for Medical Research in Kansas City, could pay for hundreds of new science and technology jobs.

But leaders of those institutions believe the state could do more. Anticipated tobacco-settlement money, which would have augmented jobs created by the funding, never materialized after the state hit a tough economic patch. And more modest plans to boost science in the region, by the governments of both Missouri and neighbouring Kansas, may be limited by laws in both states: Kansas will fund stem-cell research under the same restrictions as US federal funding, whereas Missouri may ban state funding for stem-cell research entirely.

Despite these obstacles, the privately funded institutions are more than holding their own. William Neaves, president of the Stowers Institute, points to a series of successes since the \$200-million facility opened its doors in 2000 (see *Nature* **408**, 281; 2000).

He has recruited 20 principal investigators, including a few who gave up prestigious Howard Hughes Medical Institute awards to join. He even convinced developmental biologist Olivier Pourquié to move from sunny Marseilles, France, to slightly less glamorous Kansas City — where Pourquié, however, quickly located a pastry shop that met his standards.

CELL DIVISIONS

The Stowers has made a splash scientifically too, including high-impact stem-cell papers by Pourquié and Julien Dubrulle (*Nature* **427**, 419–422; 2004) and by Linheng Li's group (J. Zhang *et al.* *Nature* **425**, 836–841; 2003). Buoyed by these successes, the institute announced in February that it would build another facility to hold 20 principal investigators.

But there are some clouds. The state governments of Kansas and Missouri, which Neaves had hoped would help the region by funnelling money into local institutions such as the University of Missouri-Kansas City and the University of Kansas Medical Center, have not invested “as rapidly as we’d have hoped”, Neaves says.

And money that the states do provide may come with limitations, owing to restrictive stem-cell policies. Before the Stowers announcement in February, rumours abounded that if Missouri did not ease its anti-stem-cell stance, the Stowers might locate the new building in another state, one more receptive to stem-cell

research. The donors, Jim and Virginia Stowers, will not allow the new building in any state that restricts research on early stem cells produced by nuclear transfer. However, Neaves remains optimistic that the building will be in Missouri.

Bill Duncan, president of the Kansas City Area Life Sciences Institute (KCALSI), says that the area is being “driven by the Stowers Institute’s investment”. The original plan of this scientific regional development agency, formed when the Stowers was created, was for KCALSI’s eight member organizations to spend \$500 million a year on research and development. “That would put us in the top 10 to 15 regions,” Duncan says.

However, research and development at KCALSI has not yet reached that level. It spent \$219 million in 2003, more than twice its 1994 total, but still less than halfway to the \$500-million goal. This was partly because the institute was expecting research to benefit from money paid by tobacco companies as part of a national health-compensation settlement. When the states faced financial difficulties, the tobacco money went to pay other bills.

Duncan is cautiously optimistic that other funding will materialize. For instance, the state of Kansas passed a Bioscience Authority Act this year, which will funnel \$500 million into the state during the next 10 years. Much of this would benefit the Kansas City area, which straddles the Kansas–Missouri border, as about 65% of Kansas’s research activity takes place there. The University of Kansas Medical Center is already slowly adding 25 senior scholar positions.

Missouri is considering similar schemes, but Duncan is concerned that “overly broad legislation” on stem-cell research could stop any moves gaining momentum. “We can’t afford to mess this up,” he says.

James Spigarelli, president of the Midwest Research Institution, says that any Kansas or Missouri law against stem-cell research could paralyse recruitment by local centres, preventing the collaborations envisaged by KCALSI. “We have an uphill battle to hire the best scientists as it is,” says Spigarelli.

Attracting people who have never lived in the middle of the country can initially be tough, says Spigarelli. In fact, the area has much to offer, on top of a low cost of living. The research infrastructure in



Mark Wrighton: setting a \$300-million research agenda for Washington University.



William Neaves: proud of Stowers’ scientific achievements, but feels the state could do more.

Solid foundations: both Washington University, below, and the Stowers Institute have built strong reputations in stem-cell research. As long as the state doesn't enact a ban, the future looks bright for further developments.



Sitting in his office, in the tower of a century-old Neo-gothic building that might have been airlifted from the University of Cambridge in England, Washington University chancellor Mark Wrighton was eager to share news that the university's medical school had just edged out Harvard's in receiving the second-highest amount of National Institutes of Health funding, trailing only Johns Hopkins University Medical School in Baltimore, Maryland.

Wrighton was also keen to tout his new research agenda, called Biomed 21, a \$300-million scheme to organize biomedical research around three units — the Center for Genomics and Human Genetics, the Center for Biological Imaging, and the Division of Clinical Sciences. The programmes will provide training for 50 PhD, MD and MD/PhD students. This effort is complemented by a fund-raising campaign that has raised \$1.5 billion, enough for 165 endowed professorships and 11 endowed faculty fellowships.

One programme was funded by former faculty member Phil Needleman — now an associate dean at Washington's medical school — who developed the blockbuster arthritis drug Celebrex while on his way to becoming chief scientific officer at Pharmacia. Wrighton believes this history hints that Pfizer will stay committed to St Louis. Although other Pharmacia sites are being closed now that Pfizer owns the company, the St Louis site has been named the company's global centre for cardiovascular and inflammatory disease.

Washington University isn't the only local recipient of philanthropy. A \$75-million facility for the not-for-profit Donald Danforth Plant Science Center opened three years ago, with funding from the St Louis-based Danforth Foundation and the Monsanto Foundation, and a tax credit from the state of Missouri. The centre now has 11 principal investigators, supported by about 190 other scientists. It plans to increase its staff to 17 principal investigators, along with their teams of students, postdocs and technicians.

The Danforth centre is also supported by Washington University. Its collaborations with Washington and several other universities, as well as the Missouri Botanical Garden, put St Louis at the heart of plant science research.

With private support for Washington University, the Stowers Institute and the Danforth Center, St Louis has shown that it can build world-class centres and attract high-calibre scientists. State investment to bolster such centres — and freedom from restrictive stem-cell laws — may be what the state needs to turn the corner and attain its goal of becoming one of the top research and development states in the country. **n**

Paul Smaglik is editor of *Naturejobs*.

Kansas City boasts facilities that rival institutions on both coasts. And the city is far from being a cultural wasteland — as attested by sculptor Claes Oldenburg's giant shuttlecocks on the lawn of the Nelson-Atkins Museum of Art, visible from Spigarelli's office window.

But restrictive laws could be ruinous, especially if Californians vote in next month's state referendum to channel millions of dollars into stem-cell research. "If our laws are more restrictive on stem cells, we're going to miss the boat for 30 years," says Spigarelli.

MEET ME IN ST LOUIS

St Louis, 250 miles to the east of Kansas City, has its own challenges. The future is looking uncertain for two of the area's biggest scientific employers: Monsanto, which has downsized over the past few years, and Pfizer, which has reduced and consolidated sites nationally since acquiring Pharmacia last year. But Washington University, like the Stowers to the west, is holding strong.

GRADUATE JOURNAL

The science of religion

Last night, my wife took me to a gospel concert commemorating the 9/11 World Trade Center tragedy. I hadn't been to church in a while, and the atmosphere moved me — almost to tears. As we left, I wondered why I had been so affected. Then I realized that it was the first time in a while that I had been in a room of people who held the same beliefs as I did. You see, I was raised as a Christian in an African-Methodist Episcopal church. I have yet to meet someone in a laboratory with the same religious background.

Most of the folks I've known in science are atheists or agnostics, so I've always been in the minority. As we live in a time of extremism, I am careful not to talk about my faith in the lab for fear of being lumped in with fundamentalists. However, my time as the lone 'believer' has made me wonder, "Is science hostile to religion?"

Religion and science are, by nature, antithetical ideals — one requiring proof, the other, faith. So it makes sense that people are often at odds over them. I still wonder, though, if my belief will work against me on the scientific career track. Will some people think I am not an objective or serious scientist because of my beliefs? If so, amen (so be it)! I will let my data speak for me, and hope that everything else will fall into place. n

Tshaka Cunningham is a fifth-year graduate student at Rockefeller University in New York.

NUTS & BOLTS

Totally stumped?

The zinger, the poser ... It's that dreaded question that seems to come out of left field in an interview. It's one thing to present your achievements, but the ultimate challenge comes in dealing with a question that you have no good answer to or would rather not have been asked. To avoid a 'rabbit in the headlights' response, anticipate potential stumpers and keep cool when they present themselves.

Everyone has their own most dreaded questions, but stumpers tend to fall into one of two categories: either totally open-ended or probing a sensitive area. Open-ended questions leave you wondering where you might possibly begin. Enquiries such as "Tell me about yourself," and "Why should we hire you?" are perfect examples. But these questions also present the perfect invitation to shine.



With Deb Koen
Careers consultant

Before heading to an interview, settle on three key points you want to make. Then respond to open-ended questions with success stories to highlight your key messages.

For those questions to which it seems there is no good answer, remember that your approach to the question is as important as the specific content of your response. "Why did you lose your last job?" may be a sensitive subject. But preparing an honest yet concise and positive explanation can help. And illegal questions, such as "Are you married?" or

"How old are you anyway?" are best met with a curious and non-defensive question in reply, such as "Why do you ask?". This lets the interviewer put underlying concerns on the table.

Anticipating difficult questions allows you to plan your responses. Once in the interview, remember that impressions are formed more by tone and non-verbal cues than by specific words. Even if you need a little time to respond, an upbeat tone and open body language will show your interest and enthusiasm.

Use the closing moments and a friendly follow-up letter to express your interest and reaffirm important considerations. Regardless of the difficulty of the question, the approach you take in the preparation, delivery and follow-up will let you shine throughout the process. n

Deb Koen is vice-president of Career Development Services and a columnist for *The Wall Street Journal's CareerJournal.com*.

MOVERS

Nouria Hernandez and Winship Herr, professors, Center for Integrative Genomics, University of Lausanne, Switzerland



If you are a couple who both work in scientific research, you probably know how hard it is to find jobs in proximity as your careers progress, and almost impossible to find such a situation where you have family roots. But Nouria Hernandez and Winship Herr have beaten the odds.

The biologists, who met at Cold Spring

Harbor Laboratory (CSHL) in New York, last month moved to a new centre in Switzerland, where Hernandez is a citizen and Herr lived as a child.

CSHL helped to set them on their path, but they agree that they had reached a point where a move was necessary. CSHL is geared towards helping scientists start a career. "It's not the kind of place you retire," Hernandez says. "You're always imagining your exit strategy," agrees Herr.

The couple stayed there longer than most because they found CSHL an exciting place to do science, and because Herr got involved with developing, then running, the Watson School of Biological Sciences, CSHL's graduate school.

Realizing that Lausanne would help them focus on their family as well as their science required a certain amount of introspection. Herr recommends that all scientists apply that to their career trajectory.

Students and postdocs should always be thinking about where they want to be in five years' time, Herr says. Mid-career scientists need to be equally forward-looking. Scientists need to realize that opportunities arise when things go well experimentally and when good publications come their way — but such moments aren't always guaranteed.

Both emphasize the importance of good mentors. Herr also notes the importance of factors beyond science in making a career decision. He gets recharged by doing science "in beautiful places". At Lausanne, he can stroll along Lake Geneva or hike in the Alps. And when he or Hernandez hit a snag, they can find each other on the same floor and commiserate, before returning to their respective labs. n



CV

Nouria Hernandez
1987–2004: Professor, Cold Spring Harbor Laboratory, New York

1983–86: Postdoc, Yale University, Connecticut

1983: PhD, German Cancer Research Center and University of Heidelberg

Winship Herr

1983–2004: Cold Spring Harbor Laboratory (postdoc to dean of the Watson School of Biological Sciences)

1982: Postdoc, Medical Research Council, Cambridge, UK

1982: PhD, Harvard University, Massachusetts